

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 138.669 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-F

Perfect score: 7276

Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARALTTRSGQTLSKVWVNC 1363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 7276 | 100.0 | 1363 | 13 | US-10-124-557-52 |
| 2 | 7261.9 | 99.8 | 1404 | 9 | Sequence 52, Appl |
| 3 | 7261.9 | 99.8 | 1404 | 11 | Sequence 30, Appl |
| 4 | 7261.9 | 99.8 | 1404 | 11 | US-09-897-188-1 |
| 5 | 7261.9 | 99.8 | 1404 | 13 | Sequence 1, Appl |
| 6 | 7058.7 | 97.0 | 1320 | 13 | US-10-124-557-2 |
| 7 | 7058.7 | 97.0 | 1320 | 13 | US-10-124-557-62 |
| 8 | 7044.6 | 96.8 | 1361 | 13 | Sequence 62, Appl |
| 9 | 7002 | 96.2 | 1313 | 13 | Sequence 46, Appl |
| 10 | 6987.9 | 96.0 | 1354 | 13 | Sequence 60, Appl |
| 11 | 6878.9 | 94.5 | 1314 | 13 | Sequence 40, Appl |
| 12 | 6794.7 | 93.4 | 1270 | 13 | Sequence 142, Appl |
| 13 | 6780.6 | 93.2 | 1311 | 13 | Sequence 48, Appl |
| 14 | 5820.9 | 80.0 | 1140 | 13 | Sequence 50, Appl |
| | | | | | Sequence 44, Appl |
| | | | | | Sequence 42, Appl |
| | | | | | Sequence 104, Appl |

| | | | | | | |
|----|--------|------|-------|----|----------------------|-------------------|
| 15 | 5561 | 76.4 | 1049 | 13 | US-10-124-557-58 | Sequence 58, Appl |
| 16 | 5495.5 | 75.5 | 1038 | 13 | US-10-124-557-74 | Sequence 74, Appl |
| 17 | 5231.9 | 71.9 | 1022 | 13 | US-10-124-557-84 | Sequence 84, Appl |
| 18 | 5011 | 68.9 | 941 | 13 | US-10-124-557-14 | Sequence 14, Appl |
| 19 | 3897.7 | 53.6 | 792 | 9 | US-09-802-207-27 | Sequence 27, Appl |
| 20 | 2850.9 | 39.2 | 538 | 14 | US-10-038-694-3 | Sequence 3, Appl |
| 21 | 2160.9 | 29.7 | 422 | 13 | US-10-124-557-68 | Sequence 68, Appl |
| 22 | 2146.8 | 29.5 | 463 | 13 | US-10-124-557-54 | Sequence 54, Appl |
| 23 | 2032.8 | 27.9 | 423 | 13 | US-10-124-557-66 | Sequence 66, Appl |
| 24 | 1886.9 | 25.9 | 372 | 13 | US-10-124-557-64 | Sequence 64, Appl |
| 25 | 1714.7 | 23.6 | 401 | 9 | US-09-802-207-29 | Sequence 29, Appl |
| 26 | 1401.6 | 19.3 | 292 | 16 | US-10-468-910-4 | Sequence 4, Appl |
| 27 | 1371 | 18.8 | 5179 | 9 | US-08-922-217-1068 | Sequence 1068, Ap |
| 28 | 1371 | 18.8 | 5179 | 9 | US-09-833-263-1068 | Sequence 1068, Ap |
| 29 | 1371 | 18.8 | 5179 | 13 | US-10-025-380-1068 | Sequence 1068, Ap |
| 30 | 1371 | 18.8 | 5179 | 16 | US-10-734-564-121 | Sequence 121, App |
| 31 | 1200.1 | 16.5 | 296 | 13 | US-10-124-557-70 | Sequence 70, Appl |
| 32 | 992.9 | 13.6 | 1460 | 14 | US-10-295-027-428 | Sequence 428, App |
| 33 | 983 | 13.5 | 1367 | 9 | US-09-801-368-108 | Sequence 108, App |
| 34 | 968.5 | 13.3 | 6642 | 14 | US-10-369-493-5013 | Sequence 5013, Ap |
| 35 | 965.1 | 13.3 | 3507 | 14 | US-10-369-493-5784 | Sequence 5784, Ap |
| 36 | 956.9 | 13.2 | 1325 | 9 | US-09-864-761-35612 | Sequence 35612, A |
| 37 | 953.7 | 13.1 | 19723 | 15 | US-10-084-846A-5 | Sequence 5, Appli |
| 38 | 946.2 | 13.0 | 237 | 13 | US-10-124-557-72 | Sequence 72, Appl |
| 39 | 941.9 | 12.9 | 5935 | 14 | US-10-243-243A-8 | Sequence 8, Appli |
| 40 | 938.6 | 12.9 | 5877 | 14 | US-10-142-515-11 | Sequence 11, Appl |
| 41 | 937.1 | 12.9 | 22152 | 16 | US-10-715-066-5 | Sequence 5, Appli |
| 42 | 923.9 | 12.7 | 220 | 13 | US-10-124-557-96 | Sequence 96, Appl |
| 43 | 912 | 12.5 | 2090 | 16 | US-10-408-765A-2318 | Sequence 2318, Ap |
| 44 | 897.3 | 12.3 | 4322 | 16 | US-10-437-963-104793 | Sequence 104793, |
| 45 | 895.7 | 12.3 | 2971 | 14 | US-10-146-473-50 | Sequence 50, Appl |

ALIGNMENTS

RESULT 1

US-10-124-557-52

; Sequence 52, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

APPLICANT: Turner, Katherine C.

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: G1 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match 100.0%; Score 7276; DB 13; Length 1363;
Best Local Similarity 100.0%; Pred. No. 1.1e-167;
Matches 1363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQELSCKGRCFESFERGRCDCDAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQELSCKGRCFESFERGRCDCDAQCKYDKCCPDYE 60

QY 61 SFCAEYHNPTSPSSKKAPPSPGASQTIKSTTKRSPKPNKKTKKVIIEEIEEHSVS 120
Db 61 SFCAEYHNPTSPSSKKAPPSPGASQTIKSTTKRSPKPNKKTKKVIIEEIEEHSVS 120

QY 121 ENCESSSSSSSSSTIWKIKSKNSAANRELQKKLVKDNKNRKTTPKPPVYDE 180
Db 121 ENCESSSSSSSSSTIWKIKSKNSAANRELQKKLVKDNKNRKTTPKPPVYDE 180

QY 181 AGSGLNDGPKVTPDSTTQHNKVSTSPKIITAKPINRPSLPNPNDSKETSILVNKE 240
Db 181 AGSGLNDGPKVTPDSTTQHNKVSTSPKIITAKPINRPSLPNPNDSKETSILVNKE 240

QY 241 TTVETKETTNNKQTSIDGKEKTSKETSIAKDLAPTSKVLAKTPKAEATTGK 300
Db 241 TTVETKETTNNKQTSIDGKEKTSKETSIAKDLAPTSKVLAKTPKAEATTGK 300

QY 301 PALTTKPEPTTPKEPASTTPKEPTTIKSAFTTPKEPAPTTKSAFTTPKEPAPTTT 360
Db 301 PALTTKPEPTTPKEPASTTPKEPTTIKSAFTTPKEPAPTTKSAFTTPKEPAPTTT 360

QY 361 KEPAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTT 420
Db 361 KEPAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTT 420

QY 421 TTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKE 480
Db 421 TTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKE 480

QY 481 PAPTTTSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTT 540
Db 481 PAPTTTSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTT 540

QY 541 KEPAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFT 600
Db 541 KEPAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFT 600

QY 601 PTTPEELAPTTPEEPTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKE 660
Db 601 PTTPEELAPTTPEEPTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKEPAPTTKSAFTTPKE 660

QY 661 KETAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTT 720
Db 661 KETAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTTKGTAPTT 720

QY 721 PKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 780
Db 721 PKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 780

QY 781 TPKETAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 840

Db 781 TPKETAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 840

QY 841 ABEPTKALENSPKPEGVPTTKTAAATKPEMTTAKDKTERDLRTTPTTTAAAPKMTKET 900
Db 841 ABEPTKALENSPKPEGVPTTKTAAATKPEMTTAKDKTERDLRTTPTTTAAAPKMTKET 900

QY 901 ATTTEKTESKITATTTQVTSITTTQDTPFKITLTKTTTLAPKVTITTKITTTTEIMNKP 960
Db 901 ATTTEKTESKITATTTQVTSITTTQDTPFKITLTKTTTLAPKVTITTKITTTTEIMNKP 960

QY 961 EETAKPKDRATNSKATTPPKQKPTKAPKPTSTPKPKTMRVRKPKTTPTRKMTSTMP 1020
Db 961 EETAKPKDRATNSKATTPPKQKPTKAPKPTSTPKPKTMRVRKPKTTPTRKMTSTMP 1020

QY 1021 LNPTSRIAEAMLOTTTRPNQTPNSKLVNPKSEDAAGAGETPHMLLRHVFMEVTPD 1080
Db 1021 LNPTSRIAEAMLOTTTRPNQTPNSKLVNPKSEDAAGAGETPHMLLRHVFMEVTPD 1080

QY 1081 MDYLRVPVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSP 1140
Db 1081 MDYLRVPVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSP 1140

QY 1141 ARRIEHWGIPSPIDVTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKDIFKGFGLTG 1200
Db 1141 ARRIEHWGIPSPIDVTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKDIFKGFGLTG 1200

QY 1201 QIVAAALSTAKYKWPESVVFKEGSGTQOVIYKQEPVKCPCRRPALNYPVYGMTOVRR 1260
Db 1201 QIVAAALSTAKYKWPESVVFKEGSGTQOVIYKQEPVKCPCRRPALNYPVYGMTOVRR 1260

QY 1261 RRERAIKPSQTHTRIQVSPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPNIRK 1320
Db 1261 RRERAIKPSQTHTRIQVSPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPNIRK 1320

QY 1321 PDGYDYAFKSDQYINIDVPSRTARAITTRSGQTLKSVWYNCP 1363
Db 1321 PDGYDYAFKSDQYINIDVPSRTARAITTRSGQTLKSVWYNCP 1363

RESULT 2
US-09-802-207-30
Sequence 30, Application US/09802207
Publication No. US2002008624A1
GENERAL INFORMATION:
APPLICANT: Warman, Matthew
APPLICANT: Carpten, John
APPLICANT: Trent, Jeffrey
APPLICANT: Marcelino, Jose
TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
FILE REFERENCE: Case-06212
CURRENT APPLICATION NUMBER: US/09/802,207
CURRENT FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: 09/619,175
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 60/145,328
PRIOR FILING DATE: 1999-07-23
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn version 3.0
SEQ ID NO 30
LENGTH: 1404
TYPE: PRT
ORGANISM: Homo sapiens
US-09-802-207-30

Query Match 99.8%; Score 7261.9; DB 9; Length 1404;
Best Local Similarity 97.1%; Pred. No. 2.6e-167;
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGYSRDATCNCNDYNCQHYMECCPDF 60

QY 26 -----ELSCGRCFESFERGECDDAQQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
Db 61 KRVCYTAELSCGRCFESFERGECDDAQQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 139
Db 121 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 180
QY 140 KIKSKNSAANRELOKLVKNDKNKRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 199
Db 181 KIKSKNSAANRELOKLVKNDKNKRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 240
QY 200 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQTSIDG 259
Db 241 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQTSIDG 300
QY 260 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTTKEPAS 319
Db 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTTKEPAS 360
QY 320 TTPKEPTPTTITKSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 379
Db 361 TTPKEPTPTTITKSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
QY 380 APTTTKSAPTTPKEPAPTTTTPKAPPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTPK 439
Db 421 APTTTKSAPTTPKEPAPTTTTPKAPPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTPK 480
QY 440 EPAPTAPKKAPPTTPKEPAPTTTTPKEPAPTTTTPKSPPTTPKEPAPTTTTPKEPAPTTTKEPAPT 499
Db 481 EPAPTAPKKAPPTTPKEPAPTTTTPKEPAPTTTTPKSPPTTPKEPAPTTTTPKEPAPTTTKEPAPT 540
QY 500 TTKSAPTTPKEPAPTTTTPKEPAPTTTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPTTTTPK 559
Db 541 TTKSAPTTPKEPAPTTTTPKEPAPTTTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPTTTTPK 600
QY 560 APTAPKEPAPTTPKETAPTTTPKXLTPTTPEKLAFTTPEKAPTTTPELAPTTPEEPPTT 619
Db 601 APTAPKEPAPTTPKETAPTTTPKXLTPTTPEKLAFTTPEKAPTTTPELAPTTPEEPPTT 660
QY 620 PEPAETTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEP 679
Db 661 PEPAETTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEP 720
QY 680 APTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 739
Db 721 APTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTPKG 780
QY 740 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPAPTTTPKETAPTTTPKEPAPTTTPK 799
Db 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPAPTTTPKETAPTTTPKEPAPTTTPK 840
QY 800 KPAPTTPETPPPTTSVSTPTTKEPTTHKSPDESTPELSAPTPKALENSPKEGVP 859
Db 841 KPAPTTPETPPPTTSVSTPTTKEPTTHKSPDESTPELSAPTPKALENSPKEGVP 900
QY 860 TKTPAATKEMTTTAKDKTTERDLRTTPTTTAAAPKWTETATTTTETKTTATTTTQV 919
Db 901 TKTPAATKEMTTTAKDKTTERDLRTTPTTTAAAPKWTETATTTTETKTTATTTTQV 960
QY 920 TSTTTQDTPPKITLTKTTLLAPKVTITTKKITTITEIMNKPBEATKPKDRATNSKATTPK 979
Db 961 TSTTTQDTPPKITLTKTTLLAPKVTITTKKITTITEIMNKPBEATKPKDRATNSKATTPK 1020
QY 980 POKETPAKKPTSTKPKTMPVRKPKTTTTPRKMSTTWPMLNPTSRIAEAMLQTTTRN 1039
Db 1021 POKETPAKKPTSTKPKTMPVRKPKTTTTPRKMSTTWPMLNPTSRIAEAMLQTTTRN 1080
QY 1040 QTPNSKLVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYLPVFNQGIINPMLS 1099
Db 1081 QTPNSKLVNPKSEDAGGAEGETPHMLLRPHVEMPEVTPDMDYLPVFNQGIINPMLS 1140
QY 1100 DETNICNGKPVLDGLTTLRNGTLVAFRGHYFWMLSPFSPARRITEVWGIPSPIDTFT 1159

Db 1141 DETNICNGKPVLDGLTTLRNGTLVAFRGHYFWMLSPFSPARRITEVWGIPSPIDTFT 1200
QY 1160 RCNCEGKTFKFDQYWRFTNDIKDAGYKPKPIFKGFGGLTGOIIVAALSTAKYKNWPESVY 1219
Db 1201 RCNCEGKTFKFDQYWRFTNDIKDAGYKPKPIFKGFGGLTGOIIVAALSTAKYKNWPESVY 1260
QY 1220 FFKRGGSIIQYIYKOEVPQKCPGRRPALNYPVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1279
Db 1261 FFKRGGSIIQYIYKOEVPQKCPGRRPALNYPVYVYVYVYVYVYVYVYVYVYVYVYVY 1320
QY 1280 SPARLAYODKGVLEHNEVKVLSILNRGLNPNVTSALSLNIRKPDGYDYAFSKDQYVYNDV 1339
Db 1321 SPARLAYODKGVLEHNEVKVLSILNRGLNPNVTSALSLNIRKPDGYDYAFSKDQYVYNDV 1380
QY 1340 PSRTARAITTRSGQTLSSKVMWYVNC 1363
Db 1381 PSRTARAITTRSGQTLSSKVMWYVNC 1404

RESULT 3

US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/296,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1

Query Match 99.8%; Score 7261.9; DB 11; Length 1404;
Best Local Similarity 97.1%; Pred. No. 2.6e-167;
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ-----25
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGYSRDATCNCYHCQHYMECCPDF 60

QY 26 -----ELSCGRCFESFERGECDDAQQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79

Db 61 KRVCYTAELSCGRCFESFERGECDDAQQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120

QY 80 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 139

Db 121 PPSGASQTIKSTTKRSKPKPNKKTKKVIIESEITEHHSVSENQESSSSSSSSSTI 180

QY 140 KIKSKNSAANRELOKLVKNDKNKRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 199

Db 181 KIKSKNSAANRELOKLVKNDKNKRTKKKPTPKPPVVDGAGSLDNGDFKVTPTDST 240

QY 200 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQTSIDG 259

Db 241 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQTSIDG 300

QY 260 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTTKEPAS 319

Db 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTTKEPAS 360

QY 320 TTPKEPTPTTITKSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 379

Db 361 TTPKEPTPTTITKSAPTTPKEPAPTTTTSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 420

QY 380 APTTTKSAPTTTPKEPAPTTTKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 439
Db 421 APTTTKSAPTTTPKEPAPTTTKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480
QY 440 EPAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 499
Db 481 EPAPTAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 540
QY 500 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 559
Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 600
QY 560 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 619
Db 601 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 660
QY 620 PEPAPTTTPKAAAPNTPEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 679
Db 661 PEPAPTTTPKAAAPNTPEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 720
QY 680 APTTPKPAKELAPTTTKEPTSTSDKPAPTTTPKGAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 739
Db 721 APTTPKPAKELAPTTTKEPTSTSDKPAPTTTPKGAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 780
QY 740 TAPTTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 799
Db 781 TAPTTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 840
QY 800 KPAPTTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKPEPGVPT 859
Db 841 KPAPTTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELSAETTPKALENSKPEPGVPT 900
QY 860 TKTPAATKPEMTTAKDKTLTERDLRTTPTTAAAPKMTKETATTTKTESKITATTTQV 919
Db 901 TKTPAATKPEMTTAKDKTLTERDLRTTPTTAAAPKMTKETATTTKTESKITATTTQV 960
QY 920 TSITTTQTTTPFKITTLTKTTLAPKVTTTKTITTEIMNKEEETAKPKDRATNSKATTPK 979
Db 961 TSITTTQTTTPFKITTLTKTTLAPKVTTTKTITTEIMNKEEETAKPKDRATNSKATTPK 1020
QY 980 POKTKAPKPTSTKPKTMRVVKPTTTPRKMSTMPBELNPTSRIAEAMLTQTTTTPN 1039
Db 1021 POKTKAPKPTSTKPKTMRVVKPTTTPRKMSTMPBELNPTSRIAEAMLTQTTTTPN 1080
QY 1040 QTPNSKLVEVNPXSSEDAGGABGETPHMLLRPHVFMPEVTPDMDXLPRVFNQGIINPMLS 1099
Db 1081 QTPNSKLVEVNPXSSEDAGGABGETPHMLLRPHVFMPEVTPDMDXLPRVFNQGIINPMLS 1140
QY 1100 DETNICNGKPYDGLTTLRNGTLVAPRGHYFWMLSFSPSPARRITEVWGPSPDITVFT 1159
Db 1141 DETNICNGKPYDGLTTLRNGTLVAPRGHYFWMLSFSPSPARRITEVWGPSPDITVFT 1200
QY 1160 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKIFKGFGLTGQIIVAALSTAKYKWPESVY 1219
Db 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKIFKGFGLTGQIIVAALSTAKYKWPESVY 1260
QY 1220 FFKRGSGIQYIYKQEPVQKCPGRRPALNYPVYGMTOVRRRRRPERAIGPSQTHIRIQY 1279
Db 1261 FFKRGSGIQYIYKQEPVQKCPGRRPALNYPVYGMTOVRRRRRPERAIGPSQTHIRIQY 1320
QY 1280 SPARLAYODKGVLHNEVKVSLIWRGLPNVVTSAISLNPRIKPDGDYDYAFSKDQYINIDV 1339
Db 1321 SPARLAYODKGVLHNEVKVSLIWRGLPNVVTSAISLNPRIKPDGDYDYAFSKDQYINIDV 1380
QY 1340 PSRTARAITTRSGQTLKSVWYNCP 1363
Db 1381 PSRTARAITTRSGQTLKSVWYNCP 1404

RESULT 4
US-10-124-557-2
; Sequence 2, Application US/10124557

Publication No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Gegner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luan
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1404 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-124-557-2
Query Match 99.8%; Score 7261.9; DB 13; Length 1404;
Best Local Similarity 97.1%; Pred. No. 2.6e-167;
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVVSQ----- 25
Db 1 MAWKTLPIYLLLSVFIQVVSQDLSSCAGCGYSDATCNCYDNCQHYMECCPDF 60
QY 26 -----ELSCKRCFESFERGECDDAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 79
Db 61 KRVCTAELSCKRCFESFERGECDDAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
QY 80 PPGASQTIKSTTKRSPPNKKTKKVIIESEITEHSVENQESSSSSSSSSSSTIW 139
Db 121 PPGASQTIKSTTKRSPPNKKTKKVIIESEITEHSVENQESSSSSSSSSSSTIW 180
QY 140 KIKSKNSAANREIQKLVKDNKKNTKKKPTKPPVVDAGSLDNGDKVITPDIST 199
Db 181 KIKSKNSAANREIQKLVKDNKKNTKKKPTKPPVVDAGSLDNGDKVITPDIST 240
QY 200 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETTETTTTTNKQSTIDG 259
Db 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETTETTTTTNKQSTIDG 300

260 KEKTSIAKETOSIEKTSIAKDLAPTSKVLAKPTPKAETTTKGPALTTTKEBPTTTPKBPAS 319
Db KEKTSIAKETOSIEKTSIAKDLAPTSKVLAKPTPKAETTTKGPALTTTKEBPTTTPKBPAS 360
320 TTPKEPTPTTIKSAPTTPKBPAPTTTTSAPTTTPKBPAPTTTKEBPTTTPKBPAPTTTKEP 379
Db TTPKEPTPTTIKSAPTTPKBPAPTTTTSAPTTTPKBPAPTTTKEBPTTTPKBPAPTTTKEP 420
380 APTTTKSAPTTKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 439
Db APTTTKSAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPK 480
440 EPAPTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 499
Db EPAPTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 540
500 TTKSAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 559
Db TTKSAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBP 600
560 APTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 619
Db APTAPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTT 660
620 PBPAPTTTPKAAAPNTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAP 679
Db PBPAPTTTPKAAAPNTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAP 720
680 APTTPKKAPKELAPTTTKEPTSTTSKAPTTTPKGTAPTTTPKBPAPTTTPKBPAPTTPKG 739
Db APTTPKKAPKELAPTTTKEPTSTTSKAPTTTPKGTAPTTTPKBPAPTTTPKBPAPTTPKG 780
740 TAPTTLKEBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAP 799
Db TAPTTLKEBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAPTTTPKBPAP 840
800 KPAPTTTPPTTSVSTPTTKEPTTHKSPDESTPELSAPTPKALENSPKBPVPT 859
Db KPAPTTTPPTTSVSTPTTKEPTTHKSPDESTPELSAPTPKALENSPKBPVPT 900
860 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAPTTTEKTTESKITATTQV 919
Db TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAPTTTEKTTESKITATTQV 960
920 TSTTTQDTPPKLTTLTKTTTAPKVTTKKTIITTTIMNKPSTAKPKDRATNSKATPK 979
Db TSTTTQDTPPKLTTLTKTTTAPKVTTKKTIITTTIMNKPSTAKPKDRATNSKATPK 1020
980 POKPTKAPKPTSTTKPKTTPVRKPKTTPTRKMTSTMPKMTSTMPKMTSTMPKMTSTMP 1039
Db POKPTKAPKPTSTTKPKTTPVRKPKTTPTRKMTSTMPKMTSTMPKMTSTMPKMTSTMP 1080
1040 QTPNSKLVNPKSBDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1099
Db QTPNSKLVNPKSBDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140
1100 DETNLCNGKPVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWIGISPDTVPT 1159
Db DETNLCNGKPVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWIGISPDTVPT 1200
1160 RCNCEGKTEFFFKDSQYWRFTNDIKDAGYKPKIFKGFGLTGQIVAAALSTAKYNWPSY 1219
Db RCNCEGKTEFFFKDSQYWRFTNDIKDAGYKPKIFKGFGLTGQIVAAALSTAKYNWPSY 1260
1220 FPKRGSIOQYIKQEPVQKCFRRPALNYPVYGMTQVRRRRFERRAIGPSQTHIRIQY 1279
Db FPKRGSIOQYIKQEPVQKCFRRPALNYPVYGMTQVRRRRFERRAIGPSQTHIRIQY 1320
1280 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGDYDYAFSKDQYVNDV 1339
Db SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGDYDYAFSKDQYVNDV 1380

QY 1340 PSRTARAITTRSGOTLSKVWYNCP 1363
Db 1381 PSRTARAITTRSGOTLSKVWYNCP 1404
RESULT 5
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-124-557-62

Query Match 99.8%; Score 7261.9; DB 13; Length 1404;
Best Local Similarity 97.1%; Pred. No. 2.6e-167;
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPTLYLLLSVFIQVSSQ----- 25
Db 1 MAWKTLPTLYLLLSVFIQVSSQ----- 25
QY 26 -----ELSCGRCPESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
Db 61 KRVCCTAELSCGRCPESFERGECDCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKSPKPNKKTKKVTSEETEEHSGENQESSSSSSSSSSSTIW 139
Db 121 PPSGASQIKSTTKSPKPNKKTKKVTSEETEEHSGENQESSSSSSSSSSSTIW 180

QY 140 KIKSKNSAANRELOKLVKONKONKRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 199
DB 181 KIKSKNSAANRELOKLVKONKONKRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 200 TOHNKVSTSPKLTAKPINRSLPNSDTSKETSUTVNKETTVEKETTNNKQTSIDG 259
DB 241 TOHNKVSTSPKLTAKPINRSLPNSDTSKETSUTVNKETTVEKETTNNKQTSIDG 300
QY 260 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPAITTPKEPPTTPPKEPAS 319
DB 301 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPAITTPKEPPTTPPKEPAS 360
QY 320 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 379
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 420
QY 380 APTTTKSAPTTPKEPAPTTPKAPTTPKEPAPTTPKAPTTPKEPPTTPKEPAPTTTKP 439
DB 421 APTTTKSAPTTPKEPAPTTPKAPTTPKEPAPTTPKAPTTPKEPPTTPKEPAPTTTKP 480
QY 440 EPAPTAPKXPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTTKSAPTTTKSAPTTTKBPAPT 499
DB 481 EPAPTAPKXPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTTKSAPTTTKSAPTTTKBPAPT 540
QY 500 TTKSAPTTPKESPPTTKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTTKKAPTTTKKP 559
DB 541 TTKSAPTTPKESPPTTKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTTKKAPTTTKKP 600
QY 560 APTAPKEPAPTTPKETAPTTPKLTPITPKLAPTTPEKAPTTPELAPTTPEEPTPTT 619
DB 601 APTAPKEPAPTTPKETAPTTPKLTPITPKLAPTTPEKAPTTPELAPTTPEEPTPTT 660
QY 620 PEPAPTTPKAAPNTPKEPAPTTPKAPTTPKAPTTPKAPTTPKETAPTTPKGTAPTTLKEP 679
DB 661 PEPAPTTPKAAPNTPKEPAPTTPKAPTTPKAPTTPKAPTTPKETAPTTPKGTAPTTLKEP 720
QY 680 APTPKKAPKELAPTTTKESPTSTSDKAPTTPKGTAPTTPKEPAPTTPKAPTTPKG 739
DB 721 APTPKKAPKELAPTTTKESPTSTSDKAPTTPKGTAPTTPKEPAPTTPKAPTTPKG 780
QY 740 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSOKPAPTTPKETAPTTPKEPAPTTPK 799
DB 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSOKPAPTTPKETAPTTPKEPAPTTPK 840
QY 800 KPAPTTPETPTPTTSEVSTPTTKAPTTHKSPDESTPELSAETPKALENSKPEPGVPT 859
DB 841 KPAPTTPETPTPTTSEVSTPTTKAPTTHKSPDESTPELSAETPKALENSKPEPGVPT 900
QY 860 TKTPAATKPBMTTAKDKTTERDLRTTPETTTAPKMTKETATTTTEKTTESKITATTQV 919
DB 901 TKTPAATKPBMTTAKDKTTERDLRTTPETTTAPKMTKETATTTTEKTTESKITATTQV 960
QY 920 TSITTQDTTPFKITTLKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATPK 979
DB 961 TSITTQDTTPFKITTLKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATPK 1020
QY 980 PQKPTKAPKPTSTPKKPTKTPVRVKPXTTTPPKMTSTMPELNPTSIAEAMLOTTRPN 1039
DB 1021 PQKPTKAPKPTSTPKKPTKTPVRVKPXTTTPPKMTSTMPELNPTSIAEAMLOTTRPN 1080
QY 1040 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPRVPGIILNPMLS 1099
DB 1081 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPRVPGIILNPMLS 1140
QY 1100 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPSPARRITBVGIPSPIDTFT 1159
DB 1141 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPSPARRITBVGIPSPIDTFT 1200
QY 1160 RCNCEGKTTFFPKDSQYWRFTNDIKDAGYPKPIFKGFGGLTQIIVAAALSTAKYKWPSPV 1219
DB 1201 RCNCEGKTTFFPKDSQYWRFTNDIKDAGYPKPIFKGFGGLTQIIVAAALSTAKYKWPSPV 1260

QY 1220 FFKRGSSIQQYIYKQEPVQKCPGRRPALNYPYVYGMTOVRRRRFERRAIGESQHTTIRIQY 1279
DB 1261 FFKRGSSIQQYIYKQEPVQKCPGRRPALNYPYVYGMTOVRRRRFERRAIGESQHTTIRIQY 1320
QY 1280 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNTRKPDGDYDYAFSKDQYNNIDV 1339
DB 1321 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNTRKPDGDYDYAFSKDQYNNIDV 1380
QY 1340 PRTARAITTRSGQTLISKVWYNCP 1363
DB 1381 PRTARAITTRSGQTLISKVWYNCP 1404

RESULT 6
US-10-124-557-46
; Sequence 46, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Gesner, Rodney M.
; Hewick, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match 97.0%; Score 7058.7; DB 13; Length 1320;
Best Local Similarity 96.8%; Pred. No. 2e-162;
Matches 1320; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
QY 1 MAWKTLPIYLLILLVSVFIQQVSSQELSCKRCFESFERGECDCDAQCKKYDKCCPDYE 60
DB 1 MAWKTLPIYLLILLVSVFIQQVSSQELSCKRCFESFERGECDCDAQCKKYDKCCPDYE 60

QY 61 SFCAEVHNTPSPSSKAPPPGASQTIKSTTKRSPKPNKXKTKVIESEBITEBHSVS 120
DB 61 SFCAEVHNTPSPSSKAPPPGASQTIKSTTKRSPKPNKXKTKVIESEBITE 115
QY 121 ENQESSSSSSSSSTIWIKSKNSAANRELQKKLVKDNKKNRKTKKPTPKPPVWDE 180
DB 116 -----VKDNKKNRKTKKPTPKPPVWDE 137
QY 181 AGSLONGDFKVTTPOTSTTQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSIVNKE 240
DB 138 AGSLONGDFKVTTPDTSITQHNKVSTSPKITTAKPINRPSLPPNSDTSKETSIVNKE 197
QY 241 TVVETKTTTTNNKQSTGDKKETSABKTSIKTSKOLAPTSKVLAKTTPKAAETTKG 300
DB 198 TVVETKTTTTNNKQSTGDKKETSABKTSIKTSKOLAPTSKVLAKTTPKAAETTKG 257
QY 301 PALTTKPEPTTPKPEASTTPKEPTTTIKSAPTTPKEPATTTTKSAPTTPKEPATTTT 360
DB 258 PALTTKPEPTTPKPEASTTPKEPTTTIKSAPTTPKEPATTTTKSAPTTPKEPATTTT 317
QY 361 KEPAATTPKEPATTTTKSAPTTPKEPATTTPKKPAPTTPKEPATTTTPKEPTTP 420
DB 318 KEPAATTPKEPATTTTKSAPTTPKEPATTTPKKPAPTTPKEPATTTTPKEPTTP 377
QY 421 TTPKEPATTTKEPATTTPKGPAPTAPKAPAPTTPKEPATTTPKGPAPTTPKEPTTPKE 480
DB 378 TTPKEPATTTKEPATTTPKGPAPTAPKAPAPTTPKEPATTTPKGPAPTTPKEPTTPKE 437
QY 481 PAPTTKSAPTTKKEPATTTKSAPTTPKPSPTTKKEPATTTKEPATTTPKKPAPTTP 540
DB 438 PAPTTKSAPTTKKEPATTTKSAPTTPKPSPTTKKEPATTTKEPATTTPKKPAPTTP 497
QY 541 KEPAATTPKEPATTTTKKPAPTAPKAPAPTTPKEPATTTPKKLPTTPKEPATTTPEKPA 600
DB 498 KEPAATTPKEPATTTTKKPAPTAPKAPAPTTPKEPATTTPKKLPTTPKEPATTTPEKPA 557
QY 601 PTTPEELAPTTPPEPTTPTPKEPATTTKAAAPTTPKEPATTTPKKPAPTTPKEPATTT 660
DB 558 PTTPEELAPTTPPEPTTPTPKEPATTTKAAAPTTPKEPATTTPKKPAPTTPKEPATTT 617
QY 661 KETAPTTKGTAPTTLKEPATTTPKKPAPKELAPTTPKEPTSTTSKDPAPTTPKGTAPT 720
DB 618 KETAPTTKGTAPTTLKEPATTTPKKPAPKELAPTTPKEPTSTTSKDPAPTTPKGTAPT 677
QY 721 PKEPATTPKEPATTTPKGTAPTTLKEPATTTPKKPAPKELAPTTPKGTSTTSKDPAPT 780
DB 678 PKEPATTPKEPATTTPKGTAPTTLKEPATTTPKKPAPKELAPTTPKGTSTTSKDPAPT 737
QY 781 TPKEATTPKEPATTTPKGPAPTTPPEPTTPSEVSTPTTKGPTTIHKSPDSSTPELS 840
DB 738 TPKEATTPKEPATTTPKGPAPTTPPEPTTPSEVSTPTTKGPTTIHKSPDSSTPELS 797
QY 841 AEPTPKALENSKPGVPTTKTAAKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTKET 900
DB 798 AEPTPKALENSKPGVPTTKTAAKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTKET 857
QY 901 ATTTEKTTESKITATTQVSTSTTQDTTPFKITTLTKTTLAPKVTITTKTITTTIMNKP 960
DB 858 ATTTEKTTESKITATTQVSTSTTQDTTPFKITTLTKTTLAPKVTITTKTITTTIMNKP 917
QY 961 BETAKPKDRATNSKATTPKPKQKPTKAPKPTSTTKPKTMVPRVKPTTPPRKMTSTWPE 1020
DB 918 BETAKPKDRATNSKATTPKPKQKPTKAPKPTSTTKPKTMVPRVKPTTPPRKMTSTWPE 977
QY 1021 LNPTSRIAEAMLTQTTTRNQTNSKLVNPKSEDAGGAGETPHMLLRBHVMEVETPD 1080
DB 978 LNPTSRIAEAMLTQTTTRNQTNSKLVNPKSEDAGGAGETPHMLLRBHVMEVETPD 1037
QY 1081 MDYLPRVNOGIIINPMLSDETNICNGKPVLDGLTTLRNGTLLVAFRGHVFWMLSPPSPSP 1140
DB 1038 MDYLPRVNOGIIINPMLSDETNICNGKPVLDGLTTLRNGTLLVAFRGHVFWMLSPPSPSP 1097

QY 1141 ARRITEVWGIBSIDTFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTG 1200
DB 1098 ARRITEVWGIBSIDTFTRCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVYFFFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1260
DB 1158 QIVAAALSTAKYKNWPESVYFFFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1217
QY 1261 RRPERAIGPSQHTTIQIYSPARLAYQDKGVLHNEVKVLSILWGLPNVVTSAISLPNIRK 1320
DB 1218 RRPERAIGPSQHTTIQIYSPARLAYQDKGVLHNEVKVLSILWGLPNVVTSAISLPNIRK 1277
QY 1321 PDGYDYAFSKDOYVNDIVPSRTARALTTRSGOTLSKVWYNCP 1363
DB 1278 PDGYDYAFSKDOYVNDIVPSRTARALTTRSGOTLSKVWYNCP 1320

RESULT 7

US-10-124-557-60
; Sequence 60, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; City: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-124-557-60

Query Match 97.0%; Score 7058.7; DB 13; Length 1320;
Best Local Similarity 96.8%; Pred. No. 2e-162;
Matches 1320; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQVSSQELSCGRCFBSFERGREGCDCAOCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLSVFIQVSSQELSCGRCFBSFERGREGCDCAOCKKYDKCCPDYE 60
QY 61 SFCAEVHNPSPSSSKAPPPGASQTIKSTTKRSPKPNKKTKKVIIESEBITEHSVS 120
Db 61 SFCAEVHNPSPSSSKAPPPGASQTIKSTTKRSPKPNKKTKKVIIESEBITE 115
QY 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKONKKNRTKKKPTPKPPVUDE 180
Db 116 -----VKONKKNRTKKKPTPKPPVUDE 137
QY 181 AGSLDNGDFKVTTPDTSTTOHNVKYSTSPKIITAKPINRPSLPNSDTSKETSITVNKE 240
Db 138 AGSLDNGDFKVTTPDTSTTOHNVKYSTSPKIITAKPINRPSLPNSDTSKETSITVNKE 197
QY 241 TTVETKETTNNKQSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKG 300
Db 198 TTVETKETTNNKQSTDGKEKTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAETTTKG 257
QY 301 PALATTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTKPEAPTTT 360
Db 258 PALATTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTKPEAPTTT 317
QY 361 KEPAPTTKPEAPTTTKPEAPTTTKSAPTTTPKEPAPTTPKKPAPTTTPKEPAPTTT 420
Db 318 KEPAPTTKPEAPTTTKPEAPTTTKSAPTTTPKEPAPTTPKKPAPTTTPKEPAPTTT 377
QY 421 TTPKEPAPTTKPEAPTTTPKEPAPTA PKAPAPTTTPKEPAPTTTPKEPAPTTTKEPS 480
Db 378 TTPKEPAPTTKPEAPTTTPKEPAPTA PKAPAPTTTPKEPAPTTTPKEPAPTTTKEPS 437
QY 481 PAPTTTKSAPTTTKPEAPTTTKSAPTTTPKEPSPTTPKEPAPTTTPKEPAPTTT 540
Db 438 PAPTTTKSAPTTTKPEAPTTTKSAPTTTPKEPSPTTPKEPAPTTTPKEPAPTTT 497
QY 541 KEPAPTTKPEAPTTTKKPAPTA PKAPAPTTTPKETAPTTPKKLPTTPEKLAPTTPEKPA 600
Db 498 KEPAPTTKPEAPTTTKKPAPTA PKAPAPTTTPKETAPTTPKKLPTTPEKLAPTTPEKPA 557
QY 601 PTTPEELAPTTPEEPTTTTPEEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTT 660
Db 558 PTTPEELAPTTPEEPTTTTPEEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTT 617
QY 661 KETAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPT 720
Db 618 KETAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPT 677
QY 721 PKEPAPTTKPEAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKPAPT 780
Db 678 PKEPAPTTKPEAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKPAPT 737
QY 781 TPKETAPTTKPEAPTTTPKEPAPTTTPETPTTSEVSTPTTPKEPTTIHKSPDESTBELS 840
Db 738 TPKETAPTTKPEAPTTTPKEPAPTTTPETPTTSEVSTPTTPKEPTTIHKSPDESTBELS 797
QY 841 AEPTPKALENSKPEPGVPTTKTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAPKMTKET 900
Db 798 AEPTPKALENSKPEPGVPTTKTTPAATKPEMTTAKDKTTERDLRTTTPETTTAAPKMTKET 857
QY 901 ATTTEKTTESKIATTTQVSTTQDTPPTPKITLTKTTLAPKVTTKITTEIMNKP 960
Db 858 ATTTEKTTESKIATTTQVSTTQDTPPTPKITLTKTTLAPKVTTKITTEIMNKP 917
QY 961 EETAKPKDRATNSKATTPKQKPTKAPKPTSTTKPKTMPRVKPKTTPTPRKMTSTMPE 1020
Db 918 EETAKPKDRATNSKATTPKQKPTKAPKPTSTTKPKTMPRVKPKTTPTPRKMTSTMPE 977
QY 1021 LNPTSRIAEAMLQTTTRPNQTPNSKLVENVPKSDAGAGETPHMLLRPHVFWPEVTPD 1080
Db 978 LNPTSRIAEAMLQTTTRPNQTPNSKLVENVPKSDAGAGETPHMLLRPHVFWPEVTPD 1037

QY 1081 MDYLPRVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSP 1140
Db 1038 MDYLPRVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSP 1097
QY 1141 ARRITEVMGIPSPIDTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTG 1200
Db 1098 ARRITEVMGIPSPIDTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQBPVQKCPGRRPALNYPVYGMTQVRR 1260
Db 1158 QIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQBPVQKCPGRRPALNYPVYGMTQVRR 1217
QY 1261 RRFERAIGPSOHTTIRIOYSPARLAYQDKGLHNVKVSILWRGLPNVVTSAISLPINRK 1320
Db 1218 RRFERAIGPSOHTTIRIOYSPARLAYQDKGLHNVKVSILWRGLPNVVTSAISLPINRK 1277
QY 1321 PDGYDYAFSKDQYYNIDVPSRTARAITTRSGQTLISKVWYNCP 1363
Db 1278 PDGYDYAFSKDQYYNIDVPSRTARAITTRSGQTLISKVWYNCP 1320

RESULT 8
US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match 96.8%; Score 7044.6; DB 13; Length 1361;
Best Local Similarity 94.0%; Pred. No. 4.5e-162;
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSCGRCPESFERGECDCDAQCKYDKCCDDYSCFAEVNPTSPPSKKAP 79
DB 61 KRVCTAELSCGRCPESFERGECDCDAQCKYDKCCDDYSCFAEVNPTSPPSKKAP 120

QY 80 PPSGASQTIKSTTKSPKPPNKKTKKVIIEBEITEEHSVSENQSSSSSSSSSSSTIIM 139
DB 121 PPSGASQTIKSTTKSPKPPNKKTKKVIIEBEITEE----- 156

QY 140 KIKSKNSAANRELQKULVKONKKNRTKKXPTPKPPVVDVBAAGSLDNGDFKVTTPDTST 199
DB 157 -----VKDNKNRTKKXPTPKPPVVDVBAAGSLDNGDFKVTTPDTST 197

QY 200 TOHNKVESTPKITAKPINRPSLPNPSDTSKETSITONKETTVEKETTNNKQTSIDG 259
DB 198 TOHNKVESTPKITAKPINRPSLPNPSDTSKETSITONKETTVEKETTNNKQTSIDG 257

QY 260 KEKTTSAKETQSIKBTSAKOLAPTCKVLAKPTPKAETTTKGPALTTKPEPTTTPKPEPAS 319
DB 258 KEKTTSAKETQSIKBTSAKOLAPTCKVLAKPTPKAETTTKGPALTTKPEPTTTPKPEPAS 317

QY 320 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
DB 318 TTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 377

QY 380 APITTKSAPTTKPEAPTTTKKAPTTKPEAPTTKPEPTTTPKPEAPTTTKPEAPTTPK 439
DB 378 APITTKSAPTTKPEAPTTTKKAPTTKPEAPTTKPEPTTTPKPEAPTTTKPEAPTTPK 437

QY 440 EPAPTAPKPAPTTPKPEAPTTTPKPEAPTTTKPSPTTPKPEAPTTTKSAPTTTKPEAPT 499
DB 438 EPAPTAPKPAPTTPKPEAPTTTPKPEAPTTTKPSPTTPKPEAPTTTKSAPTTTKPEAPT 497

QY 500 TTKSAPTTPKPSPTTKPEAPTTPKPEAPTTPKKAPTTPKPEAPTTTPKPEAPTTTKKP 559
DB 498 TTKSAPTTPKPSPTTKPEAPTTPKPEAPTTPKKAPTTPKPEAPTTTPKPEAPTTTKKP 557

QY 560 APAPTAPKPAPTTPKETAPTTKKLTPTTPEKLAPTTPEKAPTTPEBLAPTTPEEPTPTT 619
DB 558 APAPTAPKPAPTTPKETAPTTKKLTPTTPEKLAPTTPEKAPTTPEBLAPTTPEEPTPTT 617

QY 620 PEEPAPTTPKAAAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKEP 679
DB 618 PEEPAPTTPKAAAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTLKEP 677

QY 680 APPTPKPAKELAPTTTKBPTSTTSKAPPTPKGTAPTTPKPEAPTTPKPEAPTTPKG 739
DB 678 APPTPKPAKELAPTTTKBPTSTTSKAPPTPKGTAPTTPKPEAPTTPKPEAPTTPKG 737

QY 740 TAPTLKPEAPTTPKPAKELAPTTTKGPTSTTSKAPPTPKETAPTTPKPEAPTTPK 799
DB 738 TAPTLKPEAPTTPKPAKELAPTTTKGPTSTTSKAPPTPKETAPTTPKPEAPTTPK 797

QY 800 KPAPTTPETPTTSEVSTPTTKETPTTIHKSPDESTPELSAETTPKALSNSKPEGVPT 859
DB 798 KPAPTTPETPTTSEVSTPTTKETPTTIHKSPDESTPELSAETTPKALSNSKPEGVPT 857

QY 860 TKTPAATKPEMTTAKDKTTERDLRTPTTPTTAAPKMTKETAATTEKTESKITATTQV 919
DB 858 TKTPAATKPEMTTAKDKTTERDLRTPTTPTTAAPKMTKETAATTEKTESKITATTQV 917

QY 920 TSTTTQDTTTPFKITLTKTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKATTPK 979
DB 918 TSTTTQDTTTPFKITLTKTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKATTPK 977

QY 980 POKPTKADKKSTYKKPKTMRVVRKPKTTPTRKMTSTMPELNPTSRIAEMLOTTTRN 1039
DB 978 POKPTKADKKSTYKKPKTMRVVRKPKTTPTRKMTSTMPELNPTSRIAEMLOTTTRN 1037

QY 1040 QTPNSKLVENPKSDAGGAGETPHMLLRPHVFMPEVTPMDYLPVFNQGIINPMLS 1099
DB 1038 QTPNSKLVENPKSDAGGAGETPHMLLRPHVFMPEVTPMDYLPVFNQGIINPMLS 1097

QY 1100 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTFT 1159
DB 1098 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTFT 1157

QY 1160 RCNCEGKTFKDSQWRFTNDIKDAGYPKPIFKGFGSLTGQIVAAALSTAKYKNWPESVY 1219
DB 1158 RCNCEGKTFKDSQWRFTNDIKDAGYPKPIFKGFGSLTGQIVAAALSTAKYKNWPESVY 1217

QY 1220 FFKGGSQQIYIKQEPVQKCPGRRPALNYPVYEMTOVRRRRPERAIGPSQTHIRIQY 1279
DB 1218 FFKGGSQQIYIKQEPVQKCPGRRPALNYPVYEMTOVRRRRPERAIGPSQTHIRIQY 1277

QY 1280 SPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLNPIRKPDGYDYVAFSKDQYINIDV 1339
DB 1278 SPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLNPIRKPDGYDYVAFSKDQYINIDV 1337

QY 1340 PSRTARAITTSSGOTLSKVWYNCP 1363
DB 1338 PSRTARAITTSSGOTLSKVWYNCP 1361

RESULT 9
US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Jacobs, Kenneth
; Clark, Stephen C.
; Hewick, Rodney M.
; Geshner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851

```

; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Query Match          96.2%; Score 7002; DB 13; Length 1313;
Best Local Similarity 96.3%; Pred. No. 4.5e-161;
Matches 1313; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQOVSSQELSCKRCFESFERGBECDCDAQCKYDKCCDYE 60
Db 1 MAWKTLPIYLLLSLVFVIQOVSSQELSCKRCFESFERGBECDCDAQCKYDKCCDYE 60
QY 61 SFCAEVHNPTSPSSKXAPPSSGASQTIKSTTKRSPKPNKKTKKVIKESBEITEHSVS 120
Db 61 SFCA-----EHSVS 70
QY 121 ENQSSSSSSSSSSSTIWIKSSKNSAANRELQKLVKDKNKKNRKTKKPTPKPPVVDE 180
Db 71 ENQSSSSSSSSSSSTIWIKSSKNSAANRELQKLVKDKNKKNRKTKKPTPKPPVVDE 130
QY 181 AGSGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINPRSLPNSDTSKETSITVNIKE 240
Db 131 AGSGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINPRSLPNSDTSKETSITVNIKE 190
QY 241 TVTETKETITTNKOTSDGKEKITSAKETQIEKTSAMDAPTSKVLAKPTPKAETTTKG 300
Db 191 TVTETKETITTNKOTSDGKEKITSAKETQIEKTSAMDAPTSKVLAKPTPKAETTTKG 250
QY 301 PALATPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKGPAPTTP 360
Db 251 PALATPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKGPAPTTP 310
QY 361 KEPAPTTKEPAPTTKEPAPTTKSAPTTPKEPAPTTPKPAPTTPKPAPTTPKPPT 420
Db 311 KEPAPTTKEPAPTTKEPAPTTKSAPTTPKEPAPTTPKPAPTTPKPAPTTPKPPT 370
QY 421 TTPKEPAPTTKEPAPTTKEPAPTPAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKE 480
Db 371 TTPKEPAPTTKEPAPTTKEPAPTPAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKE 430
QY 481 PAPTTTKSAPTTKEPAPTTKSAPTTPKBSPTTPKEPAPTTTPKEPAPTTPKPAPTP 540
Db 431 PAPTTTKSAPTTKEPAPTTKSAPTTPKBSPTTPKEPAPTTTPKEPAPTTPKPAPTP 490
QY 541 KEPAPTTKEPAPTTKAPAPAPKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 600
Db 491 KEPAPTTKEPAPTTKAPAPAPKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 550
QY 601 PTTPELAPTTPEPTPTTPKEPAPTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTP 660
Db 551 PTTPELAPTTPEPTPTTPKEPAPTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTP 610
QY 661 KETAPTTKGTAPTTLKBPAPTTPKKAPKELAPTTTKEPSTTSQKAPAPTPKGTAPT 720
Db 611 KETAPTTKGTAPTTLKBPAPTTPKKAPKELAPTTTKEPSTTSQKAPAPTPKGTAPT 670
QY 721 PKEPAPTTPKBPAPTTKGTAPTTLKBPAPTTPKKAPKELAPTTTKEPSTTSQKAP 780
Db 671 PKEPAPTTPKBPAPTTKGTAPTTLKBPAPTTPKKAPKELAPTTTKEPSTTSQKAP 730
QY 781 TPKETAPTTPKBPAPTTPKKAPPTTPPTTSEVSTPTTTPKEPTTIHKSDPSTBELS 840
Db 731 TPKETAPTTPKBPAPTTPKKAPPTTPPTTSEVSTPTTTPKEPTTIHKSDPSTBELS 790
QY 841 AEPTTKALENSPKESGVPTTKPAATKPEMTTAKDKTERDLRTPETTTAAPKWTKET 900
Db 791 AEPTTKALENSPKESGVPTTKPAATKPEMTTAKDKTERDLRTPETTTAAPKWTKET 850

RESULT 10
US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csery, Luann
; REGISTRATION NUMBER: 31,822
```

```

; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match          96.0%; Score 6987.9; DB 13; Length 1354;
Best Local Similarity 93.5%; Pred. No. 1e-160;
Matches 1313; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLSVFVIQVSSQ----- 25
Db 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRGEGYRDATCNCYNCQHYMECCPDF 60
QY 26 -----ELSCKGRCFESPERGECDAQCKYDKCCDDYESFCAEVHNPTSPSSKKAP 79
Db 61 KRVCIAELSCKGRCFESPERGECDAQCKYDKCCDDYESFCA----- 105
QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIEEITEHSVSENQESSSSSSSSSSSTI 139
Db 106 -----EHSVSENQESSSSSSSSSSSTI 130
QY 140 KIKSSKNSAANRELOKLVKNDKNKTKKPTPKPPVVDKAGSLONGDFKVTTPDTST 199
Db 131 KIKSSKNSAANRELOKLVKNDKNKTKKPTPKPPVVDKAGSLONGDFKVTTPDTST 190
QY 200 TOHMKVSTSPKLTAKENRPPSLPPNSDTSKETSLSVKNKETTVEKETTNNKQTSDDG 259
Db 191 TOHMKVSTSPKLTAKENRPPSLPPNSDTSKETSLSVKNKETTVEKETTNNKQTSDDG 250
QY 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 319
Db 251 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 310
QY 320 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTKPEP 379
Db 311 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTKPEP 370
QY 380 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 439
Db 371 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 430
QY 440 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 499
Db 431 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 490
QY 500 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 559
Db 491 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP 550
QY 560 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 619
Db 551 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 610
QY 620 PEPAPPTPKAAAPNTPKPAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 679
Db 611 PEPAPPTPKAAAPNTPKPAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 670
QY 680 APTTPKPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPKG 739
Db 671 APTTPKPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPKG 730
QY 740 TAPTTLKEPAPTTTPKPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 799
Db 731 TAPTTLKEPAPTTTPKPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 790
```

```

QY 800 KPAPTTTPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKKEGVT 859
Db 791 KPAPTTTPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKKEGVT 850
QY 860 TKTPAATKPEMTTAKDKTTRDLRTTPTTETTAAPKMTKETATTTETTESKITATTI 919
Db 851 TKTPAATKPEMTTAKDKTTRDLRTTPTTETTAAPKMTKETATTTETTESKITATTI 910
QY 920 TSTTTQDTPPKIITLTKTTILAPKVTITTKITITTEIMNKPEETAAPKDRATNSKATPK 979
Db 911 TSTTTQDTPPKIITLTKTTILAPKVTITTKITITTEIMNKPEETAAPKDRATNSKATPK 970
QY 980 POKTKAPKKTSTKPKTMTPRVRKPKTTPTPRKMSTMPELNPTSRIAEAMLOTTTRPN 1039
Db 971 POKTKAPKKTSTKPKTMTPRVRKPKTTPTPRKMSTMPELNPTSRIAEAMLOTTTRPN 1030
QY 1040 QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMVDLPRVFNQGIINPMLS 1099
Db 1031 QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMVDLPRVFNQGIINPMLS 1090
QY 1100 DETNICNGKPVDTGLTTLRNGTLVAFRGHYFWMLSPFSGPPSPARRITEVWGIPSPIDTVFT 1159
Db 1091 DETNICNGKPVDTGLTTLRNGTLVAFRGHYFWMLSPFSGPPSPARRITEVWGIPSPIDTVFT 1150
QY 1160 RCNCEGTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCIIVAALSTAKYKNWPESVY 1219
Db 1151 RCNCEGTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCIIVAALSTAKYKNWPESVY 1210
QY 1220 FFKRGSGSIQYIYKQEPVQKCPGRRPALNVPYVGMTOVRRRERAIKPSQTHIRIQY 1279
Db 1211 FFKRGSGSIQYIYKQEPVQKCPGRRPALNVPYVGMTOVRRRERAIKPSQTHIRIQY 1270
QY 1280 SPARLAYODKGLVHNEVKVLSILMRGLENVVTSAISLPIRKPDGYDYVAFKDOVYNDV 1339
Db 1271 SPARLAYODKGLVHNEVKVLSILMRGLENVVTSAISLPIRKPDGYDYVAFKDOVYNDV 1330
QY 1340 PSRTARAITTRSGOTLSKQWYNCP 1363
Db 1331 PSRTARAITTRSGOTLSKQWYNCP 1354
```

```

RESULT 11
US-10-124-557-50
; Sequence 50, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;             Clark, Stephen C.
;             Jacobs, Kenneth
;             Hewick, Rodney M.
;             Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
;             STREET: 87 CambridgePark Drive
;             CITY: Cambridge
;             STATE: Massachusetts
;             COUNTRY: U.S.A.
;             ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
```

FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

Query Match 94.5%; Score 6878.9; DB 13; Length 1314;
Best Local Similarity 94.7%; Pred. No. 4.3e-158;
Matches 1292; Conservative 7; Mismatches 14; Indels 51; Gaps 2;

Qy 1 MAWKTLPIYLLLSVFIQOVSSQEL-SCKGRCFESFERGECDCDAQCKKYDKCCPDY 59
Db 1 MAWKTLPIYLLLSVFIQOVSSQDLSCACGEGYSRDATCNCYDNCQHYMECCPDF 60
Qy 60 ESFCAEVHNPTSPSSKAPPPSGASQTIKSTTKRSPPNKKTKKVIIESEBITEHSV 119
Db 61 KRVC-----TAHSV 70
Qy 120 SENGESSSSSSSSSTIWKSSKNSAANRELQKLKVNDKNKNTKKPTPKPPVVD 179
Db 71 SENGESSSSSSSSSTIWKSSKNSAANRELQKLKVNDKNKNTKKPTPKPPVVD 130
Qy 180 EAGGLNDGDKVTPDTSITQHNVKSTSPKITTAKPINRPSLPNSDTSKETSLVNK 239
Db 131 EAGGLNDGDKVTPDTSITQHNVKSTSPKITTAKPINRPSLPNSDTSKETSLVNK 190
Qy 240 ETTVETKETTNTKQTSIDGKEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTK 299
Db 191 ETTVETKETTNTKQTSIDGKEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTK 250
Qy 300 GPALTTPKEPTTPPKEPASTTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTT 359
Db 251 GPALTTPKEPTTPPKEPASTTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTT 310
Qy 360 TKPEAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTTKKPAPTTKPEAPTTKPEPT 419
Db 311 TKPEAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTTKKPAPTTKPEAPTTKPEPT 370
Qy 420 PTTKPEAPTTKPEAPTTTKPEAPTTAPKPAPTTKPEAPTTTKPEAPTTTKPEAPTT 479
Db 371 PTTKPEAPTTKPEAPTTTKPEAPTTAPKPAPTTKPEAPTTTKPEAPTTTKPEAPTT 430
Qy 480 EPAPTTKSAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAPTT 539
Db 431 EPAPTTKSAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKPAPTT 490
Qy 540 PKPEAPTTKPEAPTTTKKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEK 599
Db 491 PKPEAPTTKPEAPTTTKKPAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEK 550
Qy 600 APTTPEELAPTTPEPTPTTPBEAPTTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTT 659
Db 551 APTTPEELAPTTPEPTPTTPBEAPTTTPKAAAPNTPKPEAPTTTKPEAPTTTKPEAPTT 610
Qy 660 PKETAPTTKGTAPTTLKPEAPTTPKKPAKELAPTTTKPEPTSTSDKPAPTTKGTAPT 719
Db 611 PKETAPTTKGTAPTTLKPEAPTTPKKPAKELAPTTTKPEPTSTSDKPAPTTKGTAPT 670

Qy 720 TPKEAPTTKPEAPTTPKGTAPTTLKPEAPTTPKKPAKELAPTTTKGPTSTTSKDPAP 779
Db 671 TPKEAPTTKPEAPTTPKGTAPTTLKPEAPTTPKKPAKELAPTTTKGPTSTTSKDPAP 730
Qy 780 TTPKETAPTTKPEAPTTPKKPAPTTPEPTTSEVSTPTTKEPTTIHKSPDESPEL 839
Db 731 TTPKETAPTTKPEAPTTPKKPAPTTPEPTTSEVSTPTTKEPTTIHKSPDESPEL 790
Qy 840 SAEPKALENSKEPGVPTTKTAAKTPKEMTTAKDKTTERDLRTTPTTAAKPMKTK 899
Db 791 SAEPKALENSKEPGVPTTKTAAKTPKEMTTAKDKTTERDLRTTPTTAAKPMKTK 850
Qy 900 TATTTEKTESKITATTQVTSITTTQDTTPFKITTTTTLAPKVTTKTKITTTTEIMNK 959
Db 851 TATTTEKTESKITATTQVTSITTTQDTTPFKITTTTTLAPKVTTKTKITTTTEIMNK 910
Qy 960 PEETAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMPRVRKPKTTPPRKMTSTMP 1019
Db 911 PEETAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMPRVRKPKTTPPRKMTSTMP 970
Qy 1020 ELNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTP 1079
Db 971 ELNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTP 1030
Qy 1080 DMDYLPVPNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPPS 1139
Db 1031 DMDYLPVPNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPPS 1090
Qy 1140 PARRITVWGIPIPIDVTRCNCCEKTPFKDSQYWRFTNDIKDAGYKPKIFKGFGLT 1199
Db 1091 PARRITVWGIPIPIDVTRCNCCEKTPFKDSQYWRFTNDIKDAGYKPKIFKGFGLT 1150
Qy 1200 GOIVAAALSTAKYKNWPESVVFYKRGSGSIQQYIYKQEPVQKCPGRRPALNYPVYGMTOVR 1259
Db 1151 GOIVAAALSTAKYKNWPESVVFYKRGSGSIQQYIYKQEPVQKCPGRRPALNYPVYGMTOVR 1210
Qy 1260 RRRFERAIGESQHTIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIR 1319
Db 1211 RRRFERAIGESQHTIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIR 1270
Qy 1320 KPDGYDYAFSKDQYINIDVPSRTARALTTRSGOTLSKVVNCP 1363
Db 1271 KPDGYDYAFSKDQYINIDVPSRTARALTTRSGOTLSKVVNCP 1314

RESULT 12
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Geener, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/643,502
 FILING DATE: 18-JAN-1991
 APPLICATION NUMBER: US 07/546,114
 FILING DATE: 29-JUN-1990
 APPLICATION NUMBER: US 07/457,196
 FILING DATE: 29-DEC-1989
 APPLICATION NUMBER: US 07/390,901
 FILING DATE: 08-AUG-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Cserr, Juann
 REGISTRATION NUMBER: 31,822
 REFERENCE/DOCKET NUMBER: GI 5190
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)876-1170
 TELEFAX: (617)876-5851
 INFORMATION FOR SEQ ID NO: 44:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1270 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 44:
 US-10-124-557-44

Query Match 93.4%; Score 6794.7; DB 13; Length 1270;
 Best Local Similarity 93.2%; Pred. No. 4.4e-156;
 Matches 1270; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

| | | | |
|----|-----|--|-----|
| QY | 1 | MAWKTLPIYLLLSVFVIQVSSQELSCGRCFESFERGRECDCAQCKYDKCCPDYE | 60 |
| DB | 1 | MAWKTLPIYLLLSVFVIQVSSQELSCGRCFESFERGRECDCAQCKYDKCCPDYE | 60 |
| QY | 61 | SFCAEVHNPTSPSSKXAPPPSGASQITKSTKESPPKPKKTKVIESEITEHSVS | 120 |
| DB | 61 | SFCAE----- | 65 |
| QY | 121 | ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLKVKDKNKKRTKKKPTPKPPVVDE | 180 |
| DB | 66 | -----VKDNKKNRKKKPTPKPPVVDE | 87 |
| QY | 181 | AGSGLDGDFKVTPTDSTTQHNKVVSTSPKITTAKPINRPSLPNSDTSKETSITVNKE | 240 |
| DB | 88 | AGSGLDGDFKVTPTDSTTQHNKVVSTSPKITTAKPINRPSLPNSDTSKETSITVNKE | 147 |
| QY | 241 | TTVETKETTITNKQTSIDGKEKTSABETOSIEKTSKOLAPTSKVLAKETPKAETTKG | 300 |
| DB | 148 | TTVETKETTITNKQTSIDGKEKTSABETOSIEKTSKOLAPTSKVLAKETPKAETTKG | 207 |
| QY | 301 | PALTTKPEPTTTPKPEPASTTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTT | 360 |
| DB | 208 | PALTTKPEPTTTPKPEPASTTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTT | 267 |
| QY | 361 | KEPAPTTKPEPAPTTTKAPAPTTKGAAPTTKPEPAPTTPKKAPTTKPEPAPTTKPEPTP | 420 |
| DB | 268 | KEPAPTTKPEPAPTTTKAPAPTTKGAAPTTKPEPAPTTPKKAPTTKPEPAPTTKPEPTP | 327 |
| QY | 421 | TTPKPEPAPTTKPEPAPTTKAPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPTPKPE | 480 |
| DB | 328 | TTPKPEPAPTTKPEPAPTTKAPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPTPKPE | 480 |
| QY | 481 | PAPTTTKSAPTTTKPEPAPTTTKSAPTTTKPEPSTTTTKPEPAPTTTKPEPAPTTKPEPAPTTT | 540 |
| DB | 388 | PAPTTTKSAPTTTKPEPAPTTTKSAPTTTKPEPSTTTTKPEPAPTTTKPEPAPTTKPEPAPTTT | 447 |
| QY | 541 | KEPAPTTKPEPAPTTTKAPAPTTKAPKAPTTKPEPAPTTPKKLTPTTPEKLAPTTPEKPA | 600 |
| DB | 448 | KEPAPTTKPEPAPTTTKAPAPTTKAPKAPTTKPEPAPTTPKKLTPTTPEKLAPTTPEKPA | 507 |
| QY | 601 | PTTPEELAPTTPEEPPTTTPPEPAPTTPKAAPNTKPEPAPTTKPEPAPTTKPEPAPTTT | 660 |
| DB | 508 | PTTPEELAPTTPEEPPTTTPPEPAPTTPKAAPNTKPEPAPTTKPEPAPTTKPEPAPTTT | 567 |

| | | | |
|----|------|--|------|
| QY | 661 | KETAPTTKPGTAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKPAPTTPKGTAPTT | 720 |
| DB | 568 | KETAPTTKPGTAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKPAPTTPKGTAPTT | 627 |
| QY | 721 | PKAPAPTTPKBPAPTTTKGAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKPAPT | 780 |
| DB | 628 | PKAPAPTTPKBPAPTTTKGAPTTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKPAPT | 687 |
| QY | 781 | TPKETAPTTPKAPAPTTTPKAPAPTTTPPTTSEVSTPTTKEPTTTHKSDESTPELS | 840 |
| DB | 688 | TPKETAPTTPKAPAPTTTPKAPAPTTTPPTTSEVSTPTTKEPTTTHKSDESTPELS | 747 |
| QY | 841 | AEPTPKALENSPKBPAPTTTKAPATKPEMTTAKDKTTERDLRTTPTTAAAPKWTKET | 900 |
| DB | 748 | AEPTPKALENSPKBPAPTTTKAPATKPEMTTAKDKTTERDLRTTPTTAAAPKWTKET | 807 |
| QY | 901 | ATTTEKTTESKITATTTQVSTTTTODTTPPKITLTKTTTLAPKVTTTKTTTTEIMNKP | 960 |
| DB | 808 | ATTTEKTTESKITATTTQVSTTTTODTTPPKITLTKTTTLAPKVTTTKTTTTEIMNKP | 867 |
| QY | 961 | EETAKPKDRATNSKATTPKQKTPKAPKPTSTTKPKTMPRVKPKTTPTPRKMSTMPKE | 1020 |
| DB | 868 | EETAKPKDRATNSKATTPKQKTPKAPKPTSTTKPKTMPRVKPKTTPTPRKMSTMPKE | 927 |
| QY | 1021 | LNPTSRITAEAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPD | 1080 |
| DB | 928 | LNPTSRITAEAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPD | 987 |
| QY | 1081 | MDYLPRVNPNGIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFSPSP | 1140 |
| DB | 988 | MDYLPRVNPNGIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFSPSP | 1047 |
| QY | 1141 | ARRITEVWGIPIPIDVFTTRCNCEGKTFKKDSQYWRFTNDIKDAGYKPKIFKGGGLTG | 1200 |
| DB | 1048 | ARRITEVWGIPIPIDVFTTRCNCEGKTFKKDSQYWRFTNDIKDAGYKPKIFKGGGLTG | 1107 |
| QY | 1201 | QIVAAALSTAKYKNWPESVYFFKRGSGSQIYVQEPVQKCPGRRPALNYPVYGMTQVRR | 1260 |
| DB | 1108 | QIVAAALSTAKYKNWPESVYFFKRGSGSQIYVQEPVQKCPGRRPALNYPVYGMTQVRR | 1167 |
| QY | 1261 | RRFERALGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSLWRGLPNVVTSAISLPNIRK | 1320 |
| DB | 1168 | RRFERALGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSLWRGLPNVVTSAISLPNIRK | 1227 |
| QY | 1321 | PDGYDYAFSKQOYINIDVPSRTARAITRSQGTLSKVWYNCP | 1363 |
| DB | 1228 | PDGYDYAFSKQOYINIDVPSRTARAITRSQGTLSKVWYNCP | 1270 |

RESULT 13
 US-10-124-557-42
 ; Sequence 42, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ;
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

| | | | | | |
|---|--|--|-----|--|--|
| CURRENT APPLICATION DATA: | | | QY | | |
| APPLICATION NUMBER: US/10/124,557 | | | 560 | | |
| FILING DATE: 16-Apr-2002 | | | Db | | |
| CLASSIFICATION: <Unknown> | | | 508 | | |
| PRIOR APPLICATION DATA: | | | QY | | |
| APPLICATION NUMBER: US 07/643,502 | | | Db | | |
| FILING DATE: 18-JAN-1991 | | | QY | | |
| APPLICATION NUMBER: US 07/546,114 | | | Db | | |
| FILING DATE: 29-JUN-1990 | | | QY | | |
| APPLICATION NUMBER: US 07/457,196 | | | Db | | |
| FILING DATE: 29-DEC-1989 | | | QY | | |
| APPLICATION NUMBER: US 07/390,901 | | | Db | | |
| FILING DATE: 08-AUG-1989 | | | QY | | |
| ATTORNEY/AGENT INFORMATION: | | | Db | | |
| NAME: Cseri, Luann | | | QY | | |
| REGISTRATION NUMBER: 31,822 | | | Db | | |
| REFERENCE/DOCKET NUMBER: GI 5190 | | | QY | | |
| TELECOMMUNICATION INFORMATION: | | | Db | | |
| TELEPHONE: (617)876-1170 | | | QY | | |
| TELEFAX: (617)876-5851 | | | Db | | |
| INFORMATION FOR SEQ ID NO: 42: | | | QY | | |
| SEQUENCE CHARACTERISTICS: | | | Db | | |
| LENGTH: 1311 amino acids | | | QY | | |
| TYPE: amino acid | | | Db | | |
| TOPOLOGY: linear | | | QY | | |
| MOLECULE TYPE: protein | | | Db | | |
| SEQUENCE DESCRIPTION: SEQ ID NO: 42: | | | QY | | |
| US-10-124-557-42 | | | Db | | |
| Query Match 93.2%; Score 6780.6; DB 13; Length 1311; | | | QY | | |
| Best Local Similarity 90.5%; Pred. No. 1e-155; | | | Db | | |
| Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2; | | | QY | | |
| 1 MAWKTLPIYLILLLSVFVIQVSSQ-----25 | | | Db | | |
| 1 MAWKTLPIYLILLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCVDYNCQHYMECCPDF 60 | | | QY | | |
| 26 -----ELSCKGRCFESPERGECDCDAQCKYDKCCPDYDFCAEVNPTSPSSKAP 79 | | | Db | | |
| 61 KRVTAEELSCKGRCFESPERGECDCDAQCKYDKCCPDYDFCAE-----106 | | | QY | | |
| 80 PPSGASQTIKTTKSPKPPNKKTKVIESEEITEHSHVSENQESSSSSSSSSSSTI 139 | | | Db | | |
| 107 -----106 | | | QY | | |
| 140 KIKSSKNSAANRELQKLVKONKKNRTKKXPTPKPPVVDVAGSLDNGDFKVTTPDTST 199 | | | Db | | |
| 107 -----VKDNKKNRTKKXPTPKPPVVDVAGSLDNGDFKVTTPDTST 147 | | | QY | | |
| 200 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKQTSDDG 259 | | | Db | | |
| 148 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKQTSDDG 207 | | | QY | | |
| 260 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTPKPEPTTTPKEPAS 319 | | | Db | | |
| 208 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTPKPEPTTTPKEPAS 267 | | | QY | | |
| 320 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPEPAPTTTKPEPTTTPKEP 379 | | | Db | | |
| 268 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPEPAPTTTKPEPTTTPKEP 327 | | | QY | | |
| 380 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKPEPAPTTTKPEPTTTPKEP 439 | | | Db | | |
| 328 APTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKPEPAPTTTKPEPTTTPKEP 387 | | | QY | | |
| 440 BPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTKPEPAPTTTKSAPTTTKPEPTTTPKEPAPTTTKPEPTTTPKEP 499 | | | Db | | |
| 388 BPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTKPEPAPTTTKSAPTTTKPEPTTTPKEPAPTTTKPEPTTTPKEP 447 | | | QY | | |
| 500 TTKSAPTTTPKEPAPTTTKPEPAPTTTPKEPAPTTTPKEPAPTTTKPEPAPTTTKPEPTTTPKEPAPTTTKPEPTTTPKEP 559 | | | Db | | |
| 448 TTKSAPTTTPKEPAPTTTKPEPAPTTTPKEPAPTTTPKEPAPTTTKPEPAPTTTKPEPTTTPKEPAPTTTKPEPTTTPKEP 507 | | | | | |

RESULT 14
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge

```

; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseixt, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
; US-10-124-557-104

Query Match      80.0%; Score 5820.9; DB 13; Length 1140;
Best Local Similarity 96.4%; Pred. No. 1.3e-132;
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFIQVVSQ-----25
Db 1 MAWKTLPIYLLLSLVFIQVVSQDLSSCAGRCGEGYSRATCNDYNCQHYMECCPDF 60
QY 26 -----ELSCGRCFESFERGECDDAQCKKYDKCCPDYEFCAEVHNPTSPSSKKAP 79
Db 61 KRVCIAELSCGRCFESFERGECDDAQCKKYDKCCPDYEFCAEVHNPTSPSSKKAP 120
QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSENQESSSSSSSSSSSTI 139
Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSENQESSSSSSSSSSSTI 180
QY 140 KIKSSXNSAANRELQKKLVKDNKNKRTKKPTPKPPVDEAGSLDNGDFKVTTPDT 199
Db 181 KIKSSXNSAANRELQKKLVKDNKNKRTKKPTPKPPVDEAGSLDNGDFKVTTPDT 240
QY 200 TQHNKYSTSPKITTAKPIPRPSLPNSDTSKETSIVNKETTVETKETTINKQSTDG 259
Db 241 TQHNKYSTSPKITTAKPIPRPSLPNSDTSKETSIVNKETTVETKETTINKQSTDG 300
QY 260 KEKTTSAKETQSTIEKTSKADLAFTSKVLAKPTPKAETTTKGPALTTPKBPTTPKEPAS 319
Db 301 KEKTTSAKETQSTIEKTSKADLAFTSKVLAKPTPKAETTTKGPALTTPKBPTTPKEPAS 360
QY 320 TTPKEPTPTTIKSAPTTPKEPARTTKSAPTTPKEPARTTKBAPTTPKEPARTTKBP 379
Db 361 TTPKEPTPTTIKSAPTTPKEPARTTKSAPTTPKEPARTTKBAPTTPKEPARTTKBP 420
QY 380 APTTTKSAPTTPKEPARTTKBAPTTPKEPARTTKBPTTPKEPARTTKBAPTTPK 439
Db 421 APTTTKSAPTTPKEPARTTKBAPTTPKEPARTTKBPTTPKEPARTTKBAPTTPK 480

```

```

QY 440 EPAPTAPKPADPTTPKEPARTTPKEPARTTPKEPARTTPKEPARTTPKEPART 499
Db 481 EPAPTAPKPADPTTPKEPARTTPKEPARTTPKEPARTTPKEPARTTPKEPART 540
QY 500 TTKSAPTTPKEPSPTTPKEPARTTPKEPARTTPKEPARTTPKEPARTTPKEPART 559
Db 541 TTKSAPTTPKEPSPTTPKEPARTTPKEPARTTPKEPARTTPKEPARTTPKEPART 600
QY 560 APTAPKEPARTTPKETAPTTPKKLTPTTPEKLAPTTPKEPARTTPKEPARTTP 619
Db 601 APTAPKEPARTTPKETAPTTPKKLTPTTPEKLAPTTPKEPARTTPKEPARTTP 660
QY 620 PEERAPTTPKAAAPNTTPKEPARTTPKEPARTTPKEPARTTPKEPARTTPKEPART 679
Db 661 PEERAPTTPKAAAPNTTPKEPARTTPKEPARTTPKEPARTTPKEPARTTPKEPART 720
QY 680 APTTPKPAKELAPTTPKPTSTTSDDKPAPTTPKGTAPTTPKGTAPTTPKGTAPT 739
Db 721 APTTPKPAKELAPTTPKPTSTTSDDKPAPTTPKGTAPTTPKGTAPTTPKGTAPT 780
QY 740 TAPTTLKEPARTTPKPAKELAPTTPKGTPTSTTSDDKPAPTTPKGTAPTTPKGTAPT 799
Db 781 TAPTTLKEPARTTPKPAKELAPTTPKGTPTSTTSDDKPAPTTPKGTAPTTPKGTAPT 840
QY 800 KPAPTTPETPPPTSEVSTPTTPKPTTIHKSPDESTPELSAETPKALENSPKPEGVPT 859
Db 841 KPAPTTPETPPPTSEVSTPTTPKPTTIHKSPDESTPELSAETPKALENSPKPEGVPT 900
QY 860 TKTPAATKPEMTTAKOKTTERDLRTTPTTAAKMTKETATTTTEKTESKITATTTQV 919
Db 901 TKTPAATKPEMTTAKOKTTERDLRTTPTTAAKMTKETATTTTEKTESKITATTTQV 960
QY 920 TSTTTQDTPPKITTLKTTTLAPKVTTTKKTTITTEIMNKDEETAKPKDRATNSKATTPK 979
Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTTKKTTITTEIMNKDEETAKPKDRATNSKATTPK 1020
QY 980 POKETKAPKPTSTKPKATMPRVKPKTTPTPRKVSTMPPELNPTSRIAEAMLOTTTRPN 1039
Db 1021 POKETKAPKPTSTKPKATMPRVKPKTTPTPRKVSTMPPELNPTSRIAEAMLOTTTRPN 1080
QY 1040 QTNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1099
Db 1081 QTNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1140

RESULT 15
US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>

```

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/643,502
;; FILING DATE: 18-JAN-1991
;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cseri, Luann
;; REGISTRATION NUMBER: 31,822
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 876-1170
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 58:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1049 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 76.4%; Score 5561; DB 13; Length 1049;
Best Local Similarity 95.5%; Pred. No. 2.2e-126;
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIQOVSSOELSCKRCFESFERGECDCDAQCKYDKCCPDYE 60
Db |||||
QY 1 MAWKTLPIYLLLSLVFVIQOVSSOELSCKRCFESFERGECDCDAQCKYDKCCPDYE 60
Db |||||
QY 61 SFCAEVHNPTGPPSSKAPPPSGASQTIKSTKSPKPPNKKTKKVIESEBEEIIEHSVS 120
Db |||||-----EHSVS 70
QY 121 ENQSSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKKRTHKKTPKPPVYDE 180
Db |||||
QY 71 ENQSSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKKRTHKKTPKPPVYDE 130
Db |||||
QY 181 AGSLDNGDFKVTTPDTSSTQHNKVSTSPKLTAKPINRPSLPNSDTSKETSIVNKE 240
Db |||||
QY 131 AGSLDNGDFKVTTPDTSSTQHNKVSTSPKLTAKPINRPSLPNSDTSKETSIVNKE 190
Db |||||
QY 241 TVVETKETITNKTSTQDGEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKG 300
Db |||||
QY 131 TVVETKETITNKTSTQDGEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKG 250
Db |||||
QY 301 PALTTPKGPTPTTPKEPASTTPKEPTPTTIKSAPTPKEPAPTTTKSAPTTKEPAPTT 360
Db |||||
QY 251 PALTTPKGPTPTTPKEPASTTPKEPTPTTIKSAPTPKEPAPTTTKSAPTTKEPAPTT 310
Db |||||
QY 361 KEPAATTPKEPAPTTTKPAPTTTKSAPTTKEPAPTTPKPAPTTTPKEPAPTTTPKEPTP 420
Db |||||
QY 311 KEPAATTPKEPAPTTTKPAPTTTKSAPTTKEPAPTTPKPAPTTTPKEPAPTTTPKEPTP 370
Db |||||
QY 421 TTPKEPAPTTKEPAPTTTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPKE 480
Db |||||
QY 371 TTPKEPAPTTKEPAPTTTPKEPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPTPKE 430
Db |||||
QY 481 PAPTPTKSAPTTKEPAPTTTKSAPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 540
Db |||||
QY 431 PAPTPTKSAPTTKEPAPTTTKSAPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 490
Db |||||
QY 541 KEPAATTPKEPAPTTTKPAPTTTKSAPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 600
Db |||||
QY 491 KEPAATTPKEPAPTTTKPAPTTTKSAPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPA 550
Db |||||
QY 601 PTPPELAPTTPEEPTPTTPKEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTT 660
Db |||||
QY 551 PTPPELAPTTPEEPTPTTPKEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTT 610
Db |||||

Search completed: October 13, 2004, 11:53:15
Job time : 145.669 secs

QY 661 KETAPTTPKGTAPTTTLKEPAPTTPKKPAKELAPTTTKBPTSTTSKPAPTTPKGTAPTT 720
Db |||||
QY 611 KETAPTTPKGTAPTTTLKEPAPTTPKKPAKELAPTTTKBPTSTTSKPAPTTPKGTAPTT 670
Db |||||
QY 721 PKEPAPTTPKEPAPTTPKGTAPTTTLKEPAPTTPKKPAKELAPTTTKBPTSTTSKPAPT 780
Db |||||
QY 671 PKEPAPTTPKEPAPTTPKGTAPTTTLKEPAPTTPKKPAKELAPTTTKBPTSTTSKPAPT 730
Db |||||
QY 781 TPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTEPELS 840
Db |||||
QY 731 TPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSTPTTTKEPTTIHKSPDESTEPELS 790
Db |||||
QY 841 AEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKET 900
Db |||||
QY 791 AEPTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKET 850
Db |||||
QY 901 ATTTEKTTESKIATTTQVTSITTTQDTPPKITTLKTTTLAPKVTTTKKTIITTEIWNKP 960
Db |||||
QY 851 ATTTEKTTESKIATTTQVTSITTTQDTPPKITTLKTTTLAPKVTTTKKTIITTEIWNKP 910
Db |||||
QY 961 EETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTTPRVRKPKTTPRKMSTMPKE 1020
Db |||||
QY 911 EETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTTPRVRKPKTTPRKMSTMPKE 970
Db |||||
QY 1021 LNPTSRIAEAMLQTTTTPNQTNPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 1080
Db |||||
QY 971 LNPTSRIAEAMLQTTTTPNQTNPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 1030
Db |||||
QY 1081 MDYLPVRVNOGIIINPMLS 1099
Db |||||
QY 1031 MDYLPVRVNOGIIINPMLS 1049
Db |||||

GenCore version 5.1.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: October 13, 2004, 11:23:49 ; Search time 149.862 Seconds
 (without alignments)
 5233.063 Million cell updates/sec

Title: SEQ1-F
 Perfect score: 7276
 Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGQTLKVVWVNC 1363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|--------------|---------------------|
| 1 | 7261.9 | 99.8 | 1404 | 2 Q92954 | Q92954 homo sapien |
| 2 | 7232.9 | 99.4 | 1404 | 2 Q9BX49 | Q9BX49 homo sapien |
| 3 | 4852 | 66.7 | 933 | 2 Q6ZM25 | Q6ZM25 homo sapien |
| 4 | 4852 | 66.7 | 933 | 2 BAD18580 | Bad18580 h cdna fl |
| 5 | 3866.9 | 53.4 | 1054 | 2 Q9JM99 | Q9JM99 mus musculus |
| 6 | 1714.7 | 23.6 | 401 | 2 Q77765 | Q77765 bos taurus |
| 7 | 1384.6 | 19.0 | 1225 | 2 Q9VR49 | Q9VR49 drosophila |
| 8 | 1371 | 18.8 | 5179 | 1 MUC2_HUMAN | Q02817 homo sapien |
| 9 | 1349.9 | 18.6 | 1761 | 2 Q7KTF6 | Q7KTF6 drosophila |
| 10 | 1349.9 | 18.6 | 1761 | 2 AAS64673 | AAS64673 drosophila |
| 11 | 1333 | 18.3 | 251 | 2 Q6DNC4 | Q6dnc4 homo sapien |
| 12 | 1324.3 | 18.2 | 3150 | 2 Q7PMD5 | Q7pmd5 anopheles g |
| 13 | 1237.4 | 17.0 | 3409 | 2 Q6SSE6 | Q6sse6 chlamydomon |
| 14 | 1237.4 | 17.0 | 3409 | 2 AAS07044 | AAS07044 chlamydom |
| 15 | 1190.4 | 16.4 | 1664 | 1 SLR1_CLOTM | Q06852 clostridium |
| 16 | 1183 | 16.3 | 251 | 2 Q6QJF3 | Q6qjf3 bos taurus |
| 17 | 1183 | 16.3 | 251 | 2 AAS83977 | AAS83977 bos taurus |
| 18 | 1172.4 | 16.1 | 3889 | 2 Q6SSE8 | Q6sse8 chlamydomon |
| 19 | 1172.4 | 16.1 | 3889 | 2 AAS07042 | AAS07042 chlamydom |
| 20 | 1154.1 | 15.9 | 9234 | 2 Q7KTP5 | Q7ktp5 drosophila |
| 21 | 1154.1 | 15.9 | 9234 | 2 AAN10531 | AAN10531 drosophila |
| 22 | 1129.6 | 15.5 | 3432 | 2 Q8IR51 | Q8ir51 drosophila |
| 23 | 1129.6 | 15.5 | 3458 | 2 Q8IR52 | Q8ir52 drosophila |
| 24 | 1117.1 | 15.4 | 1349 | 2 Q8WQ4 | Q8wwq4 homo sapien |
| 25 | 1109.2 | 15.2 | 5703 | 1 MUSB_HUMAN | Q9HC84 homo sapien |
| 26 | 1106.6 | 15.2 | 1795 | 2 Q76894 | Q76894 drosophila |
| 27 | 1105.9 | 15.2 | 1079 | 2 Q9N4S7 | Q9n4s7 caenorhabdi |
| 28 | 1103.2 | 15.2 | 23015 | 2 Q8IQ18 | Q8iq18 drosophila |
| 29 | 1103.2 | 15.2 | 23015 | 2 AAN10358 | AAN10358 drosophila |
| 30 | 1096.9 | 15.1 | 34350 | 2 Q8WZ42 | Q8wz42 homo sapien |
| 31 | 1069.7 | 14.7 | 2284 | 2 Q9VPG1 | Q9vpg1 drosophila |

| | | | | | |
|----|--------|------|-------|------------|----------------------|
| 32 | 1065.3 | 14.6 | 2187 | 2 P70670 | P70670 mus musculu |
| 33 | 1061.4 | 14.6 | 7962 | 2 Q10465 | Q10465 homo sapien |
| 34 | 1028 | 14.1 | 972 | 2 Q7QKK7 | Q7qkk7 anopheles g |
| 35 | 1027.2 | 14.1 | 10578 | 2 Q8ISF5 | Q8isf5 caenorhabdi |
| 36 | 1027.2 | 14.1 | 18519 | 2 Q8ISF6 | Q8isf6 caenorhabdi |
| 37 | 1027.2 | 14.1 | 18534 | 2 Q8ISF7 | Q8isf7 caenorhabdi |
| 38 | 1024.8 | 14.1 | 2112 | 2 Q9VEL9 | Q9vel9 drosophila |
| 39 | 1024.2 | 14.1 | 4498 | 2 Q9W2Z3 | Q9w2z3 drosophila |
| 40 | 1024 | 14.1 | 1607 | 2 Q8H6Q5 | Q8h6q5 phytophthor |
| 41 | 1024 | 14.1 | 1607 | 2 AAP74661 | AAP74661 phytophthor |
| 42 | 1023.2 | 14.1 | 1274 | 2 Q20007 | Q20007 caenorhabdi |
| 43 | 1020.2 | 14.0 | 1489 | 2 Q36449 | Q36449 phytophthor |
| 44 | 1018.6 | 14.0 | 1480 | 2 Q9LIE8 | Q9lie8 arabidopsis |
| 45 | 1018.6 | 14.0 | 3166 | 2 Q9W3Z0 | Q9w3z0 drosophila |

ALIGNMENTS

RESULT 1
 Q92954
 ID Q92954 PRELIMINARY; PRT; 1404 AA.
 AC Q92954;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Megakaryocyte stimulating factor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 SEQUENCE FROM N.A.
 RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
 RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
 RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
 RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
 RT "Purification, Biochemical Characterization, and Cloning of a Novel
 RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony
 RT Stimulating Activity.";
 RL Blood 78:279-279(1991).
 RN [2]
 SEQUENCE FROM N.A.
 RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
 RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
 RA Jacobs K., Turner K.;
 RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";
 RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
 RL Mosher D.F. (eds.);
 RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier
 RL Science Publishers B.V. (1993).
 RN [3]
 SEQUENCE FROM N.A.
 RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
 RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
 RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
 RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RL EMBL; U70136; AAB09089.1; -;
 DR HSSP; P04004; LOCO.
 DR HSP; P04004; PRG4.
 DR GO; GO:0008283; P:cell proliferation; TAS.
 DR InterPro; IPR000585; Hemopexin.
 DR InterPro; IPR001212; Somatomedin_B.
 DR Pfam; PF00045; Hemopexin; 2.
 DR Pfam; PF01033; Somatomedin_B; 2.
 DR PRINTS; PR00022; SOMATOMEDINB.
 DR SMART; SM00120; HX; 2.
 DR SMART; SM00201; SO; 2.
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
 DR PROSITE; PS00524; SOMATOMEDIN_B; 2_1.
 SQ SEQUENCE 1404 AA; 151090 MW; AAB7AD19B35F4F6 CRC64;

| | | | | | | |
|-----------------------|-----|---|---------------------|------------|--------------|--------------------|
| Query Match | | 99.8%; | Score 7261.9; | DB 2; | Length 1404; | |
| Best Local Similarity | | 97.1%; | Pred. No. 2.7e-141; | | | |
| Matches 1363; | | Conservative | 0; | Mismatches | 0; | Indels 41; Gaps 1; |
| QY | 1 | MAWKLPYVLLLLSVFVIQQVSSQ | ----- | ----- | ----- | 25 |
| Db | 1 | MAWKLPYVLLLLSVFVIQQVSSQDLSSCAGRGEGYSDATCNCYDNCQHYMECCPDF | 60 | | | |
| QY | 26 | -----ELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP | 79 | | | |
| Db | 61 | KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP | 120 | | | |
| QY | 80 | PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEHVSVENQESSSSSSSSSTIW | 139 | | | |
| Db | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEHVSVENQESSSSSSSSSTIW | 180 | | | |
| QY | 140 | KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST | 199 | | | |
| Db | 181 | KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST | 240 | | | |
| QY | 200 | TOHNKYSTSPKIITAKPINRPSLPNNSDTSKETSITVKNKETTVEKETTNNKQISTDG | 259 | | | |
| Db | 241 | TOHNKYSTSPKIITAKPINRPSLPNNSDTSKETSITVKNKETTVEKETTNNKQISTDG | 300 | | | |
| QY | 260 | KEKTSASKEQSIKTSKADLPTSKVLAKPTPKAETTTKGPALTTPEKPTPTPKEPAS | 319 | | | |
| Db | 301 | KEKTSASKEQSIKTSKADLPTSKVLAKPTPKAETTTKGPALTTPEKPTPTPKEPAS | 360 | | | |
| QY | 320 | TTPEKPTPTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKKEP | 379 | | | |
| Db | 361 | TTPEKPTPTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEP | 420 | | | |
| QY | 380 | APTPTKSAPTTPKEPAPTTTKKAPTTPKEPAPTTTKPEPTPTPKEPAPTTTKKEP | 439 | | | |
| Db | 421 | APTPTKSAPTTPKEPAPTTTKKAPTTPKEPAPTTTKPEPTPTPKEPAPTTTKKEP | 480 | | | |
| QY | 440 | EPAPTAKKAPTTPKEPAPTTTKKAPTTPKEPAPTTTKSAPTTTKSAPTTTKKEP | 499 | | | |
| Db | 481 | EPAPTAKKAPTTPKEPAPTTTKKAPTTPKEPAPTTTKSAPTTTKSAPTTTKKEP | 540 | | | |
| QY | 500 | TTKSAPTTPKEPAPTTTKKAPTTPKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKP | 559 | | | |
| Db | 541 | TTKSAPTTPKEPAPTTTKKAPTTPKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKP | 600 | | | |
| QY | 560 | APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTTPKAPAPTPEELAPTTPPEPTT | 619 | | | |
| Db | 601 | APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTTPKAPAPTPEELAPTTPPEPTT | 660 | | | |
| QY | 620 | PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP | 679 | | | |
| Db | 661 | PEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP | 720 | | | |
| QY | 680 | APTTPKAPKELAPTTTKETSTSDKAPTTPKGTAPTTPKAPAPTTPKEPAPTTPKG | 739 | | | |
| Db | 721 | APTTPKAPKELAPTTTKETSTSDKAPTTPKGTAPTTPKAPAPTTPKEPAPTTPKG | 780 | | | |
| QY | 740 | TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTPKGTAPTTPKEPAPTTPK | 799 | | | |
| Db | 781 | TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTPKGTAPTTPKEPAPTTPK | 840 | | | |
| QY | 800 | KPAPTTPPTTSEVSTPTTKETPTTHKSPDSTPELSAEPKALENPKGPGVPT | 859 | | | |
| Db | 841 | KPAPTTPPTTSEVSTPTTKETPTTHKSPDSTPELSAEPKALENPKGPGVPT | 900 | | | |
| QY | 860 | TKTBAATKEMTTTAKDKTTERDLRTPTTPTTAAAPKMTKETAITTEKTTESKITATTQV | 919 | | | |
| Db | 901 | TKTBAATKEMTTTAKDKTTERDLRTPTTPTTAAAPKMTKETAITTEKTTESKITATTQV | 960 | | | |
| QY | 920 | TSTTTQDTPPKITLTKTTTLAPKVTTKKTTITTTIMNKPETAKPKDRATNSKATTPK | 979 | | | |
| Db | 961 | TSTTTQDTPPKITLTKTTTLAPKVTTKKTTITTTIMNKPETAKPKDRATNSKATTPK | 1020 | | | |
| QY | 980 | PQKPTKAPKKPTSTKKPKTMPRVKPKTTPTPKMTSTWPELNPTSRIAEAMLQTTTREN | 1039 | | | |

| | | | | | | |
|----|------|---|------|--|--|--|
| Db | 1021 | PQKPTKAPKKPTSTKKPKTMPRVKPKTTPTPKMTSTWPELNPTSRIAEAMLQTTTREN | 1080 | | | |
| QY | 1040 | QTPNSKLVEVNPKSADAGAGETPHMLLRPHVFNPEVTPDMDYLRPVNQGIINPMLS | 1099 | | | |
| Db | 1081 | QTPNSKLVEVNPKSADAGAGETPHMLLRPHVFNPEVTPDMDYLRPVNQGIINPMLS | 1140 | | | |
| QY | 1100 | DETNTCNGKFPVGLTTLTRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSPIDTFT | 1159 | | | |
| Db | 1141 | DETNTCNGKFPVGLTTLTRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSPIDTFT | 1200 | | | |
| QY | 1160 | RCNCEKTFPFXKDSQWFRFTNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPESVY | 1219 | | | |
| Db | 1201 | RCNCEKTFPFXKDSQWFRFTNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPESVY | 1260 | | | |
| QY | 1220 | FFKRGSGIOQYIYKQEPVQKCPGRRPALNYPVYVEMTQVRRRRFERRAIGPSQHTTIRIQY | 1279 | | | |
| Db | 1261 | FFKRGSGIOQYIYKQEPVQKCPGRRPALNYPVYVEMTQVRRRRFERRAIGPSQHTTIRIQY | 1320 | | | |
| QY | 1280 | SPARLAYQDKGVLFHNEVKVSIILWRGLPNVVTSAISLPIKPKDGYDYAFSKQOYINIDV | 1339 | | | |
| Db | 1321 | SPARLAYQDKGVLFHNEVKVSIILWRGLPNVVTSAISLPIKPKDGYDYAFSKQOYINIDV | 1380 | | | |
| QY | 1340 | PSRTARAITTRSGQTLISKWYNCP | 1363 | | | |
| Db | 1381 | PSRTARAITTRSGQTLISKWYNCP | 1404 | | | |

RESULT 2

Q9BX49

ID Q9BX49 PRELIMINARY; PRT; 1404 AA.

AC Q9BX49;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE BG17416.2 (WSF: megakaryocyte stimulating factor)

GN Name=BG17416.2;

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RA WRAY P.;

| RP | SEQUENCE FROM N.A. |
| RL | Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases. |
| DR | EMBL; AL133553; CAC36090.1; -- |
| DR | HSSP; P04004; 10C0 |
| DR | InterPro; IPR000585; Hemopexin. |
| DR | InterPro; IPR001212; Somatomedin_B. |
| DR | Pfam; PF00045; Hemopexin; 2. |
| DR | Pfam; PF01033; Somatomedin B; 2. |
| DR | PRINTS; PR00022; SOMATOMEDINB. |
| DR | SMART; SM00120; HX; 2. |
| DR | SMART; SM00201; SO; 2. |
| DR | PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1. |
| DR | PROSITE; PS00524; SOMATOMEDIN B; 2. |
| SQ | SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64; |

Query Match

Best Local Similarity 99.4%; Score 7232.9; DB 2; Length 1404;

Matches 1359; Conservative 0; Mismatches 4; Indels 41; Gaps 1;

QY 1 MAWKLPYVLLLLSVFVIQQVSSQ

Db 1 MAWKLPYVLLLLSVFVIQQVSSQDLSSCAGRGEGYSDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP 79

Db 61 KRVCTAELSCGRCFESFERGECDCDAQCKYDKCCPDYEFCAEVHNPTGPPSSKKAP 120

QY 80 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEHVSVENQESSSSSSSSSTIW 139

Db 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBITEHVSVENQESSSSSSSSSTIW 180

QY 140 KIKSSKNGAANRELQKKLVKONKKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 199
Db 181 KIKSSKNGAANRELQKKLVKONKKNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDST 240
QY 200 TOHNKVSSTPKITTTAKPNINPRSLPNSDTSKETSLSITVNKETTVETKETTNTKQSTSDG 259
Db 241 TOHNKVSSTPKITTTAKPNINPRSLPNSDTSKETSLSITVNKETTVETKETTNTKQSTSDG 300
QY 260 KEKTTSAKETOSIEKTSADLAPTQKVLAKPTPKAETTTKGPALATPKPEPTTTKEPAS 319
Db 301 KEKTTSAKETOSIEKTSADLAPTQKVLAKPTPKAETTTKGPALATPKPEPTTTKEPAS 360
QY 320 TTPKEPTTTTKSAPTTPKBPAPTTPKGAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 379
Db 361 TTPKEPTTTTKSAPTTPKBPAPTTPKGAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 420
QY 380 APTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 439
Db 421 APTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 480
QY 440 EPAPTAPKPPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 499
Db 481 EPAPTAPKPPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 540
QY 500 TTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 559
Db 541 TTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 600
QY 560 APTAPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 619
Db 601 APTAPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 660
QY 620 PEPAPTTPKAAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 679
Db 661 PEPAPTTPKAAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 720
QY 680 APTTPKBPAPKELAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 739
Db 721 APTTPKBPAPKELAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 780
QY 740 TAPTTLKEPAPTTPKBPAPKELAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 799
Db 781 TAPTTLKEPAPTTPKBPAPKELAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 840
QY 800 KPAPTTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSKPEGVPT 859
Db 841 KPAPTTPPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSKPEGVPT 900
QY 860 TKTPAATKPEMTTAKDKTTERDLRTPPTTAAKPKMTKETAATTTBKTTESKITATTTOV 919
Db 901 TKTPAATKPEMTTAKDKTTERDLRTPPTTAAKPKMTKETAATTTBKTTESKITATTTOV 960
QY 920 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEBETAKPKORATNSKATTPK 979
Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEBETAKPKORATNSKATTPK 1020
QY 980 POKPTKAPKPTSTKKPKTMPRVKPKTTPTRKMTSTWPELNPSTSRIAEAMLOTTTRN 1039
Db 1021 POKPTKAPKPTSTKKPKTMPRVKPKTTPTRKMTSTWPELNPSTSRIAEAMLOTTTRN 1080
QY 1040 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099
Db 1081 QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1140
QY 1100 DETNINCNGKPVGLTTLRNGTLVARGHYFWMISPPSPSPARRITEVWGISPIDTFT 1159
Db 1141 DETNINCNGKPVGLTTLRNGTLVARGHYFWMISPPSPSPARRITEVWGISPIDTFT 1200
QY 1160 RCNCEGKTTFFKDSQVWFNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWESVY 1219
Db 1201 RCNCEGKTTFFKDSQVWFNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWESVY 1260

QY 1220 FFKRGSGIOQYIKQEPVQKCPGRPALNYPYVGEMTOVRRRFFERAIGPSQHTTIRIQY 1279
Db 1261 FFKRGSGIOQYIKQEPVQKCPGRPALNYPYVGEMTOVRRRFFERAIGPSQHTTIRIQY 1320
QY 1280 SPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRUKPDGDYDYAFSKDQYNNIDV 1339
Db 1321 SPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRUKPDGDYDYAFSKDQYNNIDV 1380
QY 1340 PSRTARAITTRSGQTLKSVWYNCP 1363
Db 1381 PSRTARAITTRSGQTLKSVWYNCP 1404
ID Q6ZMZ5 PRELIMINARY; PRT; 933 AA.
AC Q6ZMZ5;
DT 05-JUL-2004 (TREMREL. 27, Created)
DT 05-JUL-2004 (TREMREL. 27, Last sequence update)
DT 05-JUL-2004 (TREMREL. 27, Last annotation update)
DE Hypothetical protein FLJ16561.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Irie T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie T., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yanashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Watanabe M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
RA Masuho Y., Nagai K., Isogai T.,
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK131434; BADI8580.1; -;
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin B.
DR Pfam; PF00045; Hemopexin; 2.
DR PRINTS; PR01033; Somatomedin B; 1.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 1.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN B; 1.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;
Query Match 66.7%; Score 4852; DB 2; Length 933;
Best Local Similarity 68.3%; Pred. No. 6.1e-92;
Matches 931; Conservative 0; Mismatches 2; Indels 430; Gaps 1;
QY 1 MAWKTLPIYLLLLLSVFIQVSSORLSCKRCFFSFERGRCDCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLLSVFIQVSSORLSCKRCFFSFERGRCDCDAQCKKYDKCCPDYE 60
QY 61 SFCAEVHNPSTPPSKKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSYS 120
Db 61 SFCAEVHNPSTPPSKKAPPPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSYS 120
QY 121 ENQESSSSSSSSSSSTTKIKSSKNANRELQKLVKONKKNRTKKKTPKPPVVD 180
Db 121 ENQESSSSSSSSSSSTTKIKSSKNANRELQKLVKONKKNRTKKKTPKPPVVD 180
QY 181 AGSGLDNGDFKVTTPDSTTQHNKYSTSPKTTAKPINRPSLPNSDTSKETSITVKE 240
Db 181 AGSGLDNGDFKVTTPDSTTQHNKYSTSPKTTAKPINRPSLPNSDTSKETSITVKE 240
QY 241 TVETKETTNTKQSTDGKEKTTSAKETQSTIEKTSADLAPTQKVLAKPTPKAETTTK 300
Db 241 TVETKETTNTKQSTDGKEKTTSAKETQSTIEKTSADLAPTQKVLAKPTPKAETTTK 300
QY 301 PALITPKPEPTTTKEPASTTPKEPTTTPKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 360

| | | | |
|------|----|--|------|
| 790 | QY | PKBAPTTPKKPADTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAETPKALE | 849 |
| 800 | QY | | 850 |
| 810 | QY | | 860 |
| 820 | QY | | 870 |
| 830 | QY | | 880 |
| 840 | QY | | 890 |
| 850 | QY | | 900 |
| 860 | QY | | 910 |
| 870 | QY | | 920 |
| 880 | QY | | 930 |
| 890 | QY | | 940 |
| 900 | QY | | 950 |
| 910 | QY | | 960 |
| 920 | QY | | 970 |
| 930 | QY | | 980 |
| 940 | QY | | 990 |
| 950 | QY | | 1000 |
| 960 | QY | | 1010 |
| 970 | QY | | 1020 |
| 980 | QY | | 1030 |
| 990 | QY | | 1040 |
| 1000 | QY | | 1050 |
| 1010 | QY | | 1060 |
| 1020 | QY | | 1070 |
| 1030 | QY | | 1080 |
| 1040 | QY | | 1090 |
| 1050 | QY | | 1100 |
| 1060 | QY | | 1110 |
| 1070 | QY | | 1120 |
| 1080 | QY | | 1130 |
| 1090 | QY | | 1140 |
| 1100 | QY | | 1150 |
| 1110 | QY | | 1160 |
| 1120 | QY | | 1170 |
| 1130 | QY | | 1180 |
| 1140 | QY | | 1190 |
| 1150 | QY | | 1200 |
| 1160 | QY | | 1210 |
| 1170 | QY | | 1220 |
| 1180 | QY | | 1230 |
| 1190 | QY | | 1240 |
| 1200 | QY | | 1250 |
| 1210 | QY | | 1260 |
| 1220 | QY | | 1270 |
| 1230 | QY | | 1280 |
| 1240 | QY | | 1290 |
| 1250 | QY | | 1300 |
| 1260 | QY | | 1310 |
| 1270 | QY | | 1320 |
| 1280 | QY | | 1330 |
| 1290 | QY | | 1340 |
| 1300 | QY | | 1350 |
| 1310 | QY | | 1360 |
| 1320 | QY | | 1370 |
| 1330 | QY | | 1380 |
| 1340 | QY | | 1390 |
| 1350 | QY | | 1400 |
| 1360 | QY | | 1410 |
| 1370 | QY | | 1420 |
| 1380 | QY | | 1430 |
| 1390 | QY | | 1440 |
| 1400 | QY | | 1450 |
| 1410 | QY | | 1460 |
| 1420 | QY | | 1470 |
| 1430 | QY | | 1480 |
| 1440 | QY | | 1490 |
| 1450 | QY | | 1500 |
| 1460 | QY | | 1510 |
| 1470 | QY | | 1520 |
| 1480 | QY | | 1530 |
| 1490 | QY | | 1540 |
| 1500 | QY | | 1550 |
| 1510 | QY | | 1560 |
| 1520 | QY | | 1570 |
| 1530 | QY | | 1580 |
| 1540 | QY | | 1590 |
| 1550 | QY | | 1600 |
| 1560 | QY | | 1610 |
| 1570 | QY | | 1620 |
| 1580 | QY | | 1630 |
| 1590 | QY | | 1640 |
| 1600 | QY | | 1650 |
| 1610 | QY | | 1660 |
| 1620 | QY | | 1670 |
| 1630 | QY | | 1680 |
| 1640 | QY | | 1690 |
| 1650 | QY | | 1700 |
| 1660 | QY | | 1710 |
| 1670 | QY | | 1720 |
| 1680 | QY | | 1730 |
| 1690 | QY | | 1740 |
| 1700 | QY | | 1750 |
| 1710 | QY | | 1760 |
| 1720 | QY | | 1770 |
| 1730 | QY | | 1780 |
| 1740 | QY | | 1790 |
| 1750 | QY | | 1800 |
| 1760 | QY | | 1810 |
| 1770 | QY | | 1820 |
| 1780 | QY | | 1830 |
| 1790 | QY | | 1840 |
| 1800 | QY | | 1850 |
| 1810 | QY | | 1860 |
| 1820 | QY | | 1870 |
| 1830 | QY | | 1880 |
| 1840 | QY | | 1890 |
| 1850 | QY | | 1900 |
| 1860 | QY | | 1910 |
| 1870 | QY | | 1920 |
| 1880 | QY | | 1930 |
| 1890 | QY | | 1940 |
| 1900 | QY | | 1950 |
| 1910 | QY | | 1960 |
| 1920 | QY | | 1970 |

| | | | |
|-------------|--|------|---------|
| 077765 | PRELIMINARY; | PRT; | 401 AA. |
| 077765 | | | |
| 077765 | | | |
| 01-NOV-1998 | (TrEMBLrel. 08, Created) | | |
| 01-NOV-1998 | (TrEMBLrel. 08, Last sequence update) | | |
| 01-MAR-2004 | (TrEMBLrel. 26, Last annotation update) | | |
| | Superficial zone protein (Fragment). | | |
| | Bos taurus (Bovine). | | |
| | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | |
| | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; | | |
| | Bovinae; Bos. | | |
| | NCBI_taxid=9913; | | |
| | [1] | | |
| | SEQUENCE FROM N.A. | | |
| | TISSUE=Articular cartilage; | | |
| | MEDLINE=99120896; PubMed=9920774; | | |
| | Flannery C.R., Hughes C.E., Schumacher B.L., Tudor D., Aydelotte M.B., | | |
| | Kuettner K.E., Caterson B.; | | |
| | "Articular cartilage superficial zone protein (SZP) is homologous to | | |
| | megakaryocyte stimulating factor precursor and is a multifunctional | | |
| | proteoglycan with potential growth-promoting, cytoprotective, and | | |
| | lubricating properties in cartilage metabolism."; | | |
| | Biochem. Biophys. Res. Commun. 254:535-541 (1999). | | |
| | [2] | | |
| | SEQUENCE FROM N.A. | | |
| | TISSUE=Articular cartilage; | | |
| | MEDLINE=99171663; PubMed=10073655; | | |
| | Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B., | | |
| | Aydelotte M.B.; | | |
| | "Immunodetection and partial cDNA sequence of the proteoglycan, | | |

| | | | | |
|-----|---|---|------|----------|
| AC | Q9VR49 | PRELIMINARY; | PRT; | 1235 AA. |
| AD | Q9VR49; | | | |
| DDT | 01-MAY-2000 | (TREMBLrel. 13, Created) | | |
| DDT | 01-JUN-2003 | (TREMBLrel. 24, Last sequence update) | | |
| DDT | 01-MAR-2004 | (TREMBLrel. 26, Last annotation update) | | |
| DE | CG3047-PA. | | | |
| GS | Name=Sgsl; ORFNames=CG3047; | | | |
| OS | Drosophila melanogaster (Fruit fly). | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | | |
| OC | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | | |
| OC | NCBI_TaxID=7227; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RP | MEDLINE=20196006; PubMed=10731132; | | | |
| RX | ADAMS M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D. | | | |
| RA | Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R. | | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson | | | |
| RA | Sutton G.G., Wortman J.R., Yeates M.D., Zhang Q., Chen L.X., | | | |
| RA | Brandon R.C., Rogers Y.H., Blazek R.G., Champagne M., Pfeiffer B. | | | |
| RA | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor | | | |
| RA | April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin | | | |
| RA | Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley | | | |
| RA | Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov | | | |
| RA | Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P. | | | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chao | | | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | | | |

RESULT /
Q9VR49
ID Q9VR49
PRELIMINARY:
PRT: 1225 AA.

[illegible]

RN [1] —

SEQUENCE FROM N.A.
MEDLINE=20196006; PubMed=10731132;
ADAMS M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.
R.A. George R.A., Lewis S.E., Richards S., Ashburner M., Henderson
Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
R.A. Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor
R.A. Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin
R.A. Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley
R.A. Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov
R.A. Borokova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.
R.A. Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chai
R.A. Cherry J.M., Cawley S., Dahlke C., Davidson L.B., Davies P.
R.A.

Wed Oct 13 12:38:26 2004

```

Db 915 PRSTTTCTSGTT--TTPRSTTKSTCAPT-----TTTPRSTTTTSTSRPTTTTTPRSTTT 968
QY 874 AKDKTTERDLRTTPETTTAAPKMTKETATTEKTTESKITATTQVTSVTTTQDTPPFKIT 933
Db 969 T---TTSRPTTTTPRSTT--TPSTSRPTTTTPRSTT-----TTSRPTTTTTPRSTT----- 1015
QY 934 TLKTTTLAPKVVTTTKKTTTITTEIMNKPEETAKPKDRATNSKATTPKQPKTKAPK----- 988
Db 1016 --KTSICAPTITTPRSTTTT-----TSRPTTTTPRSTTTT-----TTSRPTTTTTPRSTTTP 1065
QY 989 ---KPTSTKKPKMPVRKPKTPPKMT-----STMPENLPTSRIAEMLOTTTR----- 1037
Db 1066 CTSRPTTTTPRSTTTTTPRSTTTTPRSTTTTPCPTTPSASPR-----TTPTRPCP 1118
QY 1038 -----PNOT-----PNSKLVEVNPKSEDAGAGBETPHMLLRPHV 1072
Db 1119 CHPOPPYQIYPPWSWYFNTYFNPVWPQFN----- 1149
QY 1073 FMPEVTPDMYDLPRVNPQGIINPMLSDETNICNGKXPVDGLTLRNGTLVAERGHYFWML 1132
Db 1150 ---PVVPQWQMPGPNQ----- 1167
QY 1133 SPESP---SPARRITEVWGIPSPIDIVFTRCNCCEGKTFKDSQVWRFTNDKADGYPK 1189
Db 1168 LPGYPOOLPPP-----LPSP-----QWPWS-----WPK 1190
QY 1190 P-----IPKFGGGLTGQIQAALSTAKYKNWPSVYFKRGGSIQOYIYKOE 1235
Db 1191 PVVPQWQDCENICENLLKGVK----- 1215
QY 1236 PVQKC 1240
Db 1216 LIRRC 1220

RESULT 8
MUC2 HUMAN
ID MUC2 HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";

```

```

J. Clin. Invest. 88:1005-1013(1991).
-!- FUNCTION: Coats the epithelia of the intestines, airways, and
other mucus membrane-containing organs. Thought to provide a
protective, lubricating barrier against particles and infectious
agents at mucosal surfaces.
-!- SUBUNIT: Multimeric.
-!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,
bronchus, cervix and gall bladder.
-!- PTM: All cysteine residues are involved in intrachain or
interchain disulfide bonds (By similarity).
-!- POLYMORPHISM: The number of repeats is highly polymorphic and
varies among different alleles.
-!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
-!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
-!- SIMILARITY: Contains 2 VWFC domains.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; L21998; AAB95295.1; -
EMBL; M74027; AAB59875.1; -
EMBL; M94131; AAB59163.1; -
EMBL; M94132; AAB59164.1; -
PIR; A49963; A43932.
Genew; HGNC:7512; MUC2.
MIM; 158370; -
InterPro; IPR002919; Cysrich_TIL.
InterPro; IPR006208; Cys_knotE.
InterPro; IPR006207; Cys_knot_C.
InterPro; IPR006209; EGF_like.
InterPro; IPR001007; VWFC.
InterPro; IPR001846; VWFC_D.
Pfam; PF00007; Cys_knot; 1.
Pfam; PF01826; TIL; 1.
Pfam; PF00093; WMC; 1.
Pfam; PF00094; WMC; 4.
SMART; SM00041; CT; 1.
SMART; SM00214; VWC; 2.
SMART; SM00216; VWD; 4.
PROSITE; PS01185; CTCK_1; 1.
PROSITE; PS01225; CTCK_2; 1.
PROSITE; PS00022; EGF_1; UNKNOWN_1.
PROSITE; PS01208; VWFC_1; 2.
PROSITE; PS0184; VWFC_2; 2.
Glycoptrotein; Repeat; Signal.
SIGNAL 1 20 Potential.
CHAIN 21 5179 Mucin 2.
DOMAIN 1401 1747 APPROXIMATE REPEATS.
REPEAT 1401 1416 1.
REPEAT 1417 1432 2.
REPEAT 1433 1448 3.
REPEAT 1449 1464 4.
REPEAT 1465 1471 5.
REPEAT 1472 1478 6.
REPEAT 1479 1494 7A.
REPEAT 1495 1517 7B.
REPEAT 1518 1533 8A.
REPEAT 1534 1556 8B.
REPEAT 1557 1572 9A.
REPEAT 1573 1596 9B.
REPEAT 1597 1612 10A.
REPEAT 1613 1635 10B.
REPEAT 1636 1651 11A.
REPEAT 1652 1675 11B.
REPEAT 1676 1683 12.
REPEAT 1684 1699 13.
REPEAT 1700 1715 14.

```


RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.A., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Lai Z.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RA Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=2426065; PubMed=12537568;
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RA "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RT melanogaster euchromatin genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
SEQUENCE FROM N.A.
RP MEDLINE=2426069; PubMed=12537572;
RX Mirza S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bertencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
[4]
SEQUENCE FROM N.A.
RP MEDLINE=2426070; PubMed=12537573;
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
[5]
SEQUENCE FROM N.A.
RP FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
SEQUENCE FROM N.A.
RP FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003626; AA564673.1; -.
FT NON TER 1
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16B6241B3F4 CRC64;

| | | | | |
|-----------------------|--|-------------------|-------------|--------------|
| Query Match | 18.6%; | Score 1349.9; | DB 2; | Length 1761; |
| Best Local Similarity | 25.2%; | Pred. No. 33e-19; | | |
| Matches 472; | Conservative 117; | Mismatches 400; | Indels 881; | Gaps 91; |
| QY 47 | AQCKYDKCCPDYEFCAEVHNPSPSSKKAPPPGASOTIKSTTKRSPKPNKKYKK | 106 | | |
| Db 161 | AETKK-----STHNPT-----TKKSTLRITTEPTTKSSST | 191 | | |
| QY 107 | VIESEEITEHVSVE--NQESSSSSSS-----SSSSSTIWKI-----KS | 143 | | |
| Db 192 | AKTTREPTTKRETTERTTQEPSTSKTTHHTAEPATKKTTHPTTQKSTTLRITTEPTT | 251 | | |
| QY 144 | SKNSANRELQKLVKDNKNKTKKPTPKPPVDEAGSLDNGDFKVT-----PDT | 197 | | |
| Db 252 | RKSGTA-----KTTREPTK-----RETTERTTKPEST | 279 | | |
| QY 198 | S-TQHNKVSPTKITTAKPINPRPSLPNSDSKETSLSLVNKETTV----- | 243 | | |
| Db 280 | SKTTTHTHE-----TTAEPATKKTTHPT--TQKSTLRITTEPTTKSSTARTT | 329 | | |
| QY 244 | ---ETKETTT-----TNKOTSDGKEKTTSAKETQ-----SIEKTSADLAPTS | 284 | | |
| Db 330 | TKRETTERTTQEPSTSKTTHHTAEPATKKTTHPTTQKSTTLRITTEPTTK-- | 387 | | |
| QY 285 | KVLAKPTPKAET---TTKGPALATTPKEPTTTPKPEAS--TTPKEPT----- | 328 | | |
| Db 388 | KTTREPTTKRETTERTTKEP--SISKTTTHETAEPATKKTTHPTTQKSTTLRITTEPT | 445 | | |
| QY 329 | TIKSA-----PTPKXE-----PA-----PTTKSA----- | 348 | | |
| Db 446 | TKSSTARTTREPTTKRETTERTTQEPSTSKTTHHTAEPATKKTTHPTTQKSTTLR | 505 | | |
| QY 349 | ---PTTKPEAPTTTKPEAPTTTKPEAPTTTKPE-----APTTKSA----- | 391 | | |
| Db 506 | TEEPTTKRSSTAKTTRE--PTTKRETTERTTQEPSTSKTTHHTAETAKTTT | 563 | | |
| QY 392 | KEPA-----PTTKPEAPTTTKPEAPTTTKPEPTTTPKPEA----- | 434 | | |
| Db 564 | KSTTLRITTEPTTKRSSTAKTTRE--PTTKRETTERTTQEPSTSKTTHHTAEPATKKT | 621 | | |
| QY 435 | ---PTTKPEA-----PTAPKPEAPTTTKPEAPTTTKPEAPTTTKESPT----- | 479 | | |
| Db 622 | THEPTTQKSTTLRITTEPTTKRSSTAKTTRE--PTTKRETTERTTQEPSTSKTTHHTA | 679 | | |
| QY 480 | EPATTKRSSTAKTTKEPA-----PTTKSA-----PTTKPEAPTTTKPEA----- | 520 | | |
| Db 680 | EPATKKTTHPTTQKSTTLRITTEPTTKRSSTAKTTREPTTKRETTERTTKEPTTKTTT | 739 | | |
| QY 521 | -----PTTKP-----EPA-----PTTKPEAPTTTKPEA----- | 544 | | |
| Db 740 | HKTTEPTTKKTTHPTTKKSTTLKPTTEPTTKSTTKTTTKEPTTKRKTERTTKEPTT | 799 | | |
| QY 545 | -----PTTKPEAPTTTKPEA--PTAPK-----EPA-----PTT | 571 | | |
| Db 800 | RKTTTHKTTEPTTKN-----TTTKTTHEPTTKKSTTLKPTTEPTTKKSTTLKPTTEPTT | 855 | | |
| QY 572 | PKETA-----PTTKPKLT-----PTTKPKLT-----PTT-PEKLA | 595 | | |
| Db 856 | KRETTERTTQEPSTSKTTHHTAEPATKKTTHPTTQKSTTLRITTEPTTKRSSTAKTT | 915 | | |
| QY 596 | PEKPAPTTPEELAPTTPEPT-----PTTPE-----EPA-----PTTP | 628 | | |
| Db 916 | RE---PTTKRETTERTTKEPTTKRKTTHKTTEPTTKKTTHHTAEPATKKTTHPTTEPTT | 972 | | |
| QY 629 | KAAANTPKPEAPTTTKPEAPTTTKPEA-----PTTKETA-----PTTKAG | 670 | | |
| Db 973 | KTSITTKTTR-----PTTKRKTTEKTKEPTTKRKTTHHTKTEEPTTKTTKTTTKEPTTKS | 1030 | | |
| QY 671 | TA-----PTTLKEPAPTTPKKPAKELA-----PTTKPEPTSTSDKPA----- | 710 | | |
| Db 1031 | TTLKTEPTTKRKTSTTKTTREPTTKRKTTHHTKTEEPTTKTTTKEPTTKTTHHTAEPATKKTTHPT | 1090 | | |

```

QY 711 TTPKCTA-----PTTPKEPAPTTKKEPA-----PTTPKGTAPTTLKKEPA--- 749
Db 1091 TTQKSTILRIITEEPTTRKSSAKTTRTEPTTKREITERTKTEPTTKTHKTHTEPTTK 1150
QY 750 -----PTTPK-----KPAPKELAPTTKGTSTSDKPAATTPKETAPTTKKEPA----- 794
Db 1151 TTHTPTTKSTLKPTEE---PTTRK---TSTTKTTRTEPTTKRKTERTTKPTTRKTTTH 1205
QY 795 -----PTTPK-----KP-----APTTPPTTSEVSTPTTK 823
Db 1206 KTTREPTTKTTTKTHTEPTTKSTLKPTEPTTRKTSITTKTTRTEPTTKRVTERT 1265
QY 824 EP-----TTIHKSPD-----STPELSAEPKALESPKEPVGPTTKPAATKPEMTT 873
Db 1266 EPTTKTTHKTEPTTKTKTKTTHTEPTTK--STLKTEPTTKTKTSTTKTTRPT 1324
QY 874 AKDKTERDLR-----TTBETTTAAP----- 894
Db 1325 TKRVTERTTRTEPTTKTHKTEPTTKTKTKTKTTHTEPTTKSTLKPTEPTTRKT 1384
QY 895 ---KMTKETAT---TTEKTESKITATTQVTSITTDTPFKITLTKT-----TT 939
Db 1385 STTKTTRPTTKRVTERTTRTEPTTK--THKTEPTTKTKTKTTHTEPTTKSTT 1442
QY 940 LAP-----KVTTTK-----KTIITTEIMNKP-----BEAKPKDRATNSKAT 977
Db 1443 LKTEPTTRKTSITTKTTRTEPTTKRVTERTTRTEPTTKTKTKTTEPTTKTKTKT 1502
QY 978 PKPO-----KP-----TKAPKPT----- 991
Db 1503 HEPTTKSTLKPTEPTTRKTSITTKATREPTTKRVTERTTRTEPTTKTTHKTEPT 1562
QY 992 -----STKKPKTMRVRKPTTPTPRKMTSTMPELNPTSRIAEAMLOTT--- 1035
Db 1563 TKTKTKTKTKTTHTEPTTKSTL---KTEPTTRKTSITTKTTRTEPTTRKTSITTKTADQ 1618
QY 1036 -----TRPQTNKSLVFNPKSEDAGAGETP-----HMLLRPHV 1072
Db 1619 TTKRTTAEMSTTNOEPTSVETTNSSNQSTTETTESITTEHQVHHHHHHIYH---KPAD 1675
QY 1073 FMEVTPDMDYLRVFNQGIINPMSLDEINICNGKVPDGLTLRNGTLVAFRGHYFWM 1132
Db 1676 LGSILP-----LPDLP----- 1687
QY 1133 SPFSPSPARRITEVWGIPSIDVTFRCNCEGKTFFKDSQYWRFTNDIKDAGYKPIF 1192
Db 1688 -PLPLPLP-----WP-PLPLPEI-----PLPL- 1707
QY 1193 KGGGLTGQIVAAALSTAKYKNWPESVYFFKRGGSIQQYIYKQEPVQKCPGRRPALNVPV 1252
Db 1708 -----PPLPTA-----LPPLPLPLPP----- 1725
QY 1253 GEMTQVRRRFRERAGPSQHTIRIOYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSA 1312
Db 1726 -----LPEVNLTA 1733
QY 1313 ISLPIRKED 1322
Db 1734 ISLPEISLNP 1743

```

RESULT 11

```

Q6DNC4
ID Q6DNC4 PRELIMINARY; PRT; 251 AA.
AC Q6DNC4
DT 01-OCT-2004 (Tremblrel. 28, Created)
DT 01-OCT-2004 (Tremblrel. 28, Last sequence update)
DE 01-OCT-2004 (Tremblrel. 28, Last annotation update)
GN Proteoglycan 4 (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Query Match
Best Local Similarity 18.2%; Score 1324.3; DB 2; Length 3150;
Matches 460; Conservative 135; Mismatches 481; Indels 1487; Gaps 89;

QY 69 PTSP-----SSKXAPPPSGASQIKSTTKRSPPPNKKTKKVIIEETEEHSV--- 119
Db 771 PTRPTSTTTWSSASTPEPS-----TTPGTTTPTTPTSTESTDTTMSASTSEPTTPG 826

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jones A.R., Hughes C.E., Flannery C.R., Caterson B.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY653038; AAT74746.1; -.
FT NON_TER 251
FT NON_TER 251
SQ SEQUENCE 251 AA; 28659 MW; 60B1B474B62A743B CRC64;

```

```

Query Match
Best Local Similarity 18.3%; Score 1333; DB 2; Length 251;
Matches 247; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1107 GKPVDGLTLRNGTLVAFRGHYFWMSPFSPSPARRITEVWGIPSPIDTTFTRCNCEGK 1166
Db 1 GKPVDGLTLRNGTLVAFRGHYFWMSPFSPSPARRITEVWGIPSPIDTTFTRCNCEGK 60
QY 1167 TFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFKRGG 1226
Db 61 TFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFKRGG 120
QY 1227 IQQYIYKQEPVQKCPGRRPALNVPVGEVMTQVRRFRERAGPSQHTIRIOYSPARLAY 1286
Db 121 IQQYIYKQEPVQKCPGRRPALNVPVGETTQVRRFRERAGPSQHTIRIOYSPARLAY 180
QY 1287 QDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYAFSKDQYVINDVPSRTARA 1346
Db 181 QDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPDGYDYAFSKDQYVINDVPSRTARA 240
QY 1347 ITTRSGQTLTK 1357
Db 241 ITTRSGQTLTK 251

```

RESULT 12

```

Q7PMD5
ID Q7PMD5 PRELIMINARY; PRT; 3150 AA.
AC Q7PMD5
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE ENSANGP0000004655 (Fragment).
GN Name=ENSANG00000003651;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008980; EAA13969.2; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF_recept.
DR Pfam; PF04680; OGF_r_III; 80.
FT NON_TER 1
FT NON_TER 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441CB8C839 CRC64;

```

```

Query Match
Best Local Similarity 17.9%; Score 1324.3; DB 2; Length 3150;
Matches 460; Conservative 135; Mismatches 481; Indels 1487; Gaps 89;

QY 69 PTSP-----SSKXAPPPSGASQIKSTTKRSPPPNKKTKKVIIEETEEHSV--- 119
Db 771 PTRPTSTTTWSSASTPEPS-----TTPGTTTPTTPTSTESTDTTMSASTSEPTTPG 826

```

```
QY 120 -----GENOBSSSSSS-----SSSIIWKI 141
Db 827 TTRTTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTTPRPTSTESTDITMSSAST---P 883
QY 142 KSSKNAANRELQKLLKVDKNKQNRKKKPTPKPPVVDEAGS-----183
Db 884 EPSTTSGTR-----TTPTRPTDITMSSASTPEPSTTGGTTRTPT 929
QY 184 ---GLDNGDFKVTTPDTSITQHN-----KVSTSPKITAKP 216
Db 930 STSTDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTP 989
QY 217 INP-----RPSLPPNSDTSKETSIVNKETIVETK 246
Db 990 TRPTSTESTDITMSSASTPEPSTKPGTTRTTPRPTSTESTDITMSSASTPEPSTTGGT 1049
QY 247 ETWTTNKQSTDGKETSIA-----KETQSIEKTSKADLAP---282
Db 1050 RTTPT-RPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPST 1108
QY 283 ---TSKVLAKPTPKAET-----TTKGPALITPKBPTP-----311
Db 1109 TPGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTP 1168
QY 312 -----TTPKEPAS-----TTPKEP-----325
Db 1169 EPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSS 1228
QY 326 -----TPTTIKSAPTP-----KEPA--PTTKSAPT-----352
Db 1229 ASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITM 1288
QY 353 -----TTKEPA--PTTKSAPT-----KEPA-----TTPKAPAT 374
Db 1289 SASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTD 1348
QY 375 -----TTKEPA--PTTKSAPT-----KEPA-----TTPKAPAT 405
Db 1349 TTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTST 1408
QY 406 -----TPKEPAPTTPKEPTTPKEPAPT-----TKEPAP-----TTPKE 440
Db 1409 ESTDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPR 1468
QY 441 PAPT-----APKKPAP-----TTPKEPAP-----T 460
Db 1469 PTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRT 1528
QY 461 TPKEPAPTTPKEPSTTPKEPA-----PTTKSAPT-----TKEPA--PTT 501
Db 1529 TPTPTPTETMSSASTP-EPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGT 1587
QY 502 KSAPT-----KEPSTTTPKEPAPTTPKEPAP-----529
Db 1598 RTTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPST 1647
QY 530 -----TTPKKPAP-----TTPKEPAPT-----TTPKEPAPT 554
Db 1648 PGTTRTTPRPTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASSSESTPEPS 1707
QY 555 TTKKPA-----PTAP-----KEPAPTTPKEAPTTPKKLT-----586
Db 1708 TTPPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTDIT 1767
QY 587 -----TPE-----KLAPT-----TPE-----597
Db 1768 MSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTES 1827
QY 598 -----KPAPTTPKEAPTTPKEPAPT-----TPE-----621
Db 1828 TDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPT 1887
```

```
QY 622 -----EPAPTTPKAAAPNTPKEPAP-----TTP 644
Db 1888 STSTDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTP 1947
QY 645 KKPAP-----TTPKBPAPTTPKETA-----PTTKGAPT 675
Db 1948 TRPTPTETMSSASTPEPSTMTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRT 2007
QY 676 LKEPAP-----TTPKKAPKELAPT-----TKBPTST---703
Db 2008 PTRPTSTERDITMSSAYTPPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGG 2067
QY 704 -----TSDKPAPTTPKGTAPTTPKEP-----724
Db 2068 TRTTPRPTSTDTMSSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGG 2127
QY 725 -----APTTP-----KEPAPTTPKGTAPTTPKEPAP-----750
Db 2128 TTRTTPRPTSTDTMSSMSSASTPEPSTTGGTTRTTPRPTSTDTMSSMSSASTTTP 2187
QY 751 -----TTPKKAPKELA-----762
Db 2188 GTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTDTMSSMSSASTTPE 2247
QY 763 PTTTKGPTSTSDKPAPT-----TPKE-----TAPTTPKEPAP-----795
Db 2248 PSITPGTTRTTPRPTDITMSSMSSASTPEPSTTGGTTRTTPRPTSTDTMSSMSS 2307
QY 796 -----795
Db 2308 SASTPEPSTTPELLQDQHLPIPCRLVCQPLLRSHPRHLVQELLQDQHLPIPCROQ 2367
QY 796 -----TTPKKAPAT-----804
Db 2368 PLRSHRHLVQELLQDQHLPSDITMSSASTPEPSTTGGTTRTTPRPTSTDTMSS 2427
QY 805 --TPE-----TTPPTSEVSPTTPKPTTIHKS--831
Db 2428 ASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTSTESTD 2487
QY 832 --PDESTPELSAEP-----TPKALENSKPEP-----855
Db 2488 TMSASTPEPSTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPRPTD 2547
QY 856 -----GVPTTKTP-----AATKPEMTT-----873
Db 2548 TTMSSASTPEPSTTGGTTRTTPRPTDITMSSASTPEPSTTGGTTRTTPRPTSTDT 2607
QY 874 -----AKDKTTERDULRTP-----ETTAAPKMTKE 899
Db 2608 MSSMSSASTPEPSTTGGTTRTTPRPTDITMSSMSSASTATPGTTRTTPRPTSTES 2667
QY 900 TATT-----TEKTTESKITATTTQVTSTTQDTPFKITTLKTTTLAP---942
Db 2668 TDTMSSASTPEPSTTGGTTRTTPRPTSTDTMSSMSSASTPEPSTTGGTTRTTPRPT 2727
QY 943 -----KVTTTTKTIITTEIMNKBEETAKPKDRAFN---972
Db 2728 TPTDSTMSSMSSASTTGGTTRTTPRPTSTESTDITMSSASTPEPSTTGGTTRTTPR 2787
QY 973 -----SKATTPKQ-----981
Db 2788 TPTDTMSSMSSASTPEPSTTGGTTRTTPRPTSTDTMSSMSSASTPEPSTTGGTTR 2847
QY 982 ---KPT-----KAPKPT-----STKKPKTMP---1000
Db 2848 TTPRPTPTDSTMSSMSSSVTTPGTTRTTPRPTPTDSTMSSMSSASTPEPSTTGGT 2907
QY 1001 KVRKPKTTTPRKTMTSTM-----PELNPT-----SRIAEAMLQ-----1033
Db 2908 RTTTPRPTDSTMSSMSSASTPEPSTTGGTTRTTPRPTPTDSTMSSMSSASTPEPT 2967
QY 1034 -----TTPRPNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRHVFMPEVTPDMDYLP 1085
```

```

Db 2968 TTPGTRTPTPTTDTSTMSASATPKPSSTPGTTRTP-----TP----- 3011
QY 1086 RVNQGIIINPMLSDNETNCKNPKVDGLTILRNGILVAFRGHYFWMLSPFPSPARRIT 1145
Db 3012 -----TDSMSSMSSACTPEPSTTGTGTRT-----TPTR----- 3041
QY 1146 EVMGIPSPIDVFTTRCNCCEGKTFEFKDSQYWRFTNDKAGYKPIFKGFGSLTQIYAA 1205
Db 3042 -----PTPTDTT-----MSSASTPEP----- 3057
QY 1206 LSTAKYKNWPSVFFKRGSGIQYIYKQBPVQKCGRRPALNVPYGMETQVRRRREF 1265
Db 3058 -STT----- 3060
QY 1266 AIGSQTHIRIOXPARLAYQDKGVHLNVKVSILMRGLNVNVTSAISLNRKPDGYD 1325
Db 3061 ---PGTWT---TPT-----RPTTDSMSSMSSASTPE--- 3089
QY 1326 YYAFSKQYINIDVPSRT-----ARAITRSGQTLISKV-W 1359
Db 3090 -----PSTPGTTRTPTTPATTSSSRNPSSTSW 3118

RESULT 13
Q6SSE6
ID Q6SSE6 PRELIMINARY; PRT; 3409 AA.
AC Q6SSE6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Plus agglutinin.
GN Name=SAG1;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AASG0704.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PRO1222; ATROPHIN.
DR PRINTS; PRO1218; PSTLXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 17.0%; Score 1237.4; DB 2; Length 3409;
Best Local Similarity 14.0%; Pred. No. 2.5e-16;
Matches 396; Conservative 166; Mismatches 471; Indels 1796; Gaps 79;

QY 1 MAWKTLPI-----YLLLLSV----- 16
Db 50 LAWVTIPVSDALQFPNVNVTSSAATDAFSAIYIGALLDLVNVNVPWQNCINDTRRSYD 109
QY 17 ----- 16
Db 110 APWPSRCALPAVQGYDYDEYIDATVYSGNSLRPFYSTCRYPASSDYAFLEPYMPWVD 169
QY 17 -----FVIQVSSQ-----ELSCGR-----CFE 35
Db 170 DFGIYQPVLDGDMWGWFTVNTVDSNLKWDPAWTAQGAWLGGQDPRDAVWFGTHYCSW 229
QY 36 SPERGECDDAQCKYD----- 53
Db 230 PFV---EC---SSCENYDIADPYDPDKIANGIVPAVITALDFRNASLIYYLDFGMPFAG 283
QY 54 -----KCCPD----- 58
Db 284 SLDNWWYLNAYNFIGGPIPANLPTLLPSLOHLALDHCRAITPDVIRGTASLOYGWOQYPT 343

```

```

QY 59 ---YE----- 60
Db 344 GQPYEYCSGDVGTSGDTEYVISGIMPDEWGDVAVAASDSSLPWANLRTVLSNQALYX 403
QY 61 -----SFCAEVH----- 68
Db 404 PIPEGLRSASSISSWRLOQNTLGGPLPEFAAPINSLLYLGTGLGTHTWVHDPADHTQ 463
QY 69 -----PTSPFS-----SKKAPPPSGASQI-----KST 91
Db 464 GGECLAPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 523
QY 92 TKRSPKPNKKTKKVIESEBITEE-----H 117
Db 524 PPLPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 583
QY 118 SVSNQBSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 177
Db 584 PPSDPSPASS-----VPPSPEPP- 602
QY 178 VDEAGSLDNGDFKVTFTDTSTTOHNKVSTSPKITTAKPINRPSLPSPNSDTSKETSLTV 237
Db 603 -----SPKPPSPAPPSPAPPSP- 620
QY 238 NKETTIVETKETTNNKQSTGDKETTSKAKTOSIKTSKOLAPTSKVLAKPTPKARTT 297
Db 621 -----PPSPAPPSPAPPSP- 633
QY 298 TKGPAITTPKEPTPTTPKEPAST-----TKEPPTTIKSAPTTTKEPAPTTTKKAPT 350
Db 634 ---PAPPSPQPPSPVPQPPSPVPQPPSPVPQPPSPVPQPPSPVPQPPSPVPQPPSP 690
QY 351 TPKEPAPTTTKBPAPTTPK-----EPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT- 405
Db 691 SPAPPLPPSPDEPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 750
QY 406 -----TPKEPAPTTKEPTPTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT- 448
Db 751 PSPEPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 810
QY 449 PAPTTPKEPAPTTTPKEPAPT-----TTKPSPTTPK 479
Db 811 PEPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 870
QY 480 EPAPTTTKSAPTTTKBPAPTTTKSAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAP- 529
Db 871 PPSPEPPSPAPPSLEPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 929
QY 530 -----TTPKKPAPTTKEPAPTTKEPAPTTTKKAPAPTA- 565
Db 930 PPSPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 989
QY 566 -----EPAPTTTPKETAPTTPKLTPTTPKELAPTTPE--KPAPTTPEELAPT----- 610
Db 990 SPDPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1048
QY 611 -----TPEPTPTTPKEPAPTTPKAAAPN-----TPKEPAPTTTPKEPAPTTTPK- 653
Db 1049 PPSPEPPSPAPPLPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1108
QY 654 EPAPTTTPKETAPTTPKGTAPTTL----- 677
Db 1109 SPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1168
QY 678 EPAPTTTPKKPAPKELAPTT-----TKEPTSTTSDDKAPATTP 713
Db 1169 PPAPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1228
QY 714 KGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTPKGTAPTTTPKGTAPTTTPKGTAPT- 772
Db 1229 APPSP-APPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1283
QY 773 TSDKEPAPTTKETAPTTPKEPAPTTPKKAPATTPPTPPTT-TSEVSTPTTTTKEPTTIHKS 831

```



```

QY 59 ---YE----- 60
Db 344 QOPYEYCSGDVGTSGTEYVISGMIPOEWGDAVAAASDSSLPWANLRTVFLSNQALYG 403
QY 61 -----SFCAEVHN----- 68
Db 404 PIPEGLRSASSISSWRLOQNTLGCPLPEFAAPINSLLYLGTGLGTHTWVHDPADHTQ 463
QY 69 -----PTSPPS-----SKKAPPPSGASQTI-----KST 91
Db 464 GGECLAPPPSPSPSPRPPRPPPLPPSPVPPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 523
QY 92 TKRSPKPNKKTKKVISEBELTE-----H 117
Db 524 PPLPPSPSPTPVARCIOVGICDSPSPMPSPSPRPPQPPSPPPPPPPPPPPPPPPPPPP 583
QY 118 SVSNOESSSSSSSSSSSTTIWKIKSKNSAANRELQKKLVKUNKNRTKKTKTPKPPV 177
Db 584 PPSDPSPPASS-----VPPSPSPSP----- 602
QY 178 VDEAGSLDNGDFKVTTPDTSTTOHNVSTSPKITTAKPINRPSLPPNSDTSKETSITV 237
Db 603 -----SPKPPSPAPSPAPSP----- 620
QY 238 NKETTVEKTTTTNKOTSDGKEKTSKAKTQSIKTSKDLAPTSKVLAKTPKAEIT 297
Db 621 -----PPSPAPSPAPSP----- 633
QY 298 TKGPALTTKPEPTTTPKEPAST-----TPKEPTTTIKSAPTTTKEPAPTTTKGAPT 350
Db 634 ---PAPSPQPPSPVPPQPPSPVPPSKPPSPAPSPVPPSPAPSPAPSPAPSPAPSPAPP 690
QY 351 TPKEPAPTTTKEPAPTTK-----EPAPTTTKEPAPTTKGAPTTKEPAPTTKPKAPAT- 405
Db 691 SPAPPLPPSPSPAPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 750
QY 406 -----TPKEPAPTTKPEPTTTPKGPAPTTKGPAPTTKGPAPTTKGPAPTTKGPAPTT 448
Db 751 PSPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 810
QY 449 PAPTTPKEPAPTTKGPAPT-----TTKEPSPSTTPK 479
Db 811 PEPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 870
QY 480 EPAPTTKGAPTTTPKAPTTTTSAPTTPKPSPTTTPKAPTTTPKAPTP----- 529
Db 871 PPSPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 929
QY 530 -----TTPKKPAPTTKGPAPTTKGPAPTTTTPKGPAPTTTTPKGPAPTTTTPKGPAPTT 565
Db 930 PPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 969
QY 566 -----EPAPTTKAPATTPPKLTPTTPKELAPTTPE--KPAPTTPEELAPT----- 610
Db 990 SPDPSAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1048
QY 611 -----TPEPTTTPKEPAPTTPKAAAPN-----TPKEPAPTTKGPAPTTTPK----- 653
Db 1049 PPSPEPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1108
QY 654 EPAPTTKAPATTPKGTAPTTIL-----K 677
Db 1109 SPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1168
QY 678 EPAPTTPKKPAKELAPTT-----TKEPSTTSOKPAPTTTP 713
Db 1169 PPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1228
QY 714 KGTAPTTKAPATTPKGPAPTTKGTAPTTILKEPAPTTTPKGPAPKELAPTT--TKGPTST 772
Db 1229 APPSP-APSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1283
QY 773 TSDKAPATTPKAPATTPKGPAPTTTPKGPAPTTTPETPPPT-TSEVSTPTTTTKEPTTHKS 831

```

```

Db 1284 APPSPAPSP-----ADPSPEPSPAPPPPPSPVPPSPAPSPPTPPAPAPAAALPPLPPSPA 1340
QY 832 PDESTPELSABPTPKALE-----NSPKPGV---PTTKTP----- 863
Db 1341 PPLPVPSPAPSPSPPLPPPOPTPAMPPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1400
QY 864 -----AATKPE-----MTTAA----- 874
Db 1401 LPPDCTLLAQAALLSIPDAANSVFVWSAGLFTSVAPSTPPELLASFCTVCSCQLTATIALS 1460
QY 875 ----- 874
Db 1461 LVGSSRGNNYNGSSGNGYNGSDAAIORIPAGNSTDRGSSSGSGSSSGSGSGSGTETA 1520
QY 875 -----KDKTTBERDL-----RTTP----- 887
Db 1521 AEMAVDAVDQDTYQLQLSIGGVYTRVTVVVDRTTPSVGNVTLSANRIKQEPSAVGEASL 1580
QY 888 -----ETTTAAPKVT-----KETAT 902
Db 1581 NALGSKOAMLLITISFSEVPAPFDPAAASLIVTGALVAEWAADKMTFVVLAMTLPAELVA 1640
QY 903 TTEKTTESKIT----- 913
Db 1641 TAAGSSSGTSRSGNGNGNTAAAAAAPPAGTTGRRRALOQOAAAAPPSPSSSSSL 1700
QY 914 ----- 914
Db 1701 SGAATANQQQRHVHFLPATAYADAARNPGRNDLSLSVELTDNAVASPAVGEALATTARV 1760
QY 920 TSITTTQDTPPKITTTTLTKTTLPKVT----- 945
Db 1761 TAATYP-----AVAATTLVAASSSFAQAIKAGSLLQGSVHIQMLTWSLYLASRGV 1813
QY 946 -----TTTKTITITE-----IMNK 959
Db 1814 GREYGEVAFKXAVIGVKNLGPABAAAMPTNEKEVTAABQARQVGDLPIGNDLLGGS 1873
QY 960 PEETAKPKORATNSKATTPKPKP----- 983
Db 1874 SNTTSSSGSSSSSSNSP-PRPPPPPPAGSTGLLFSNADASPPPLAVATPALPAPLPS 1932
QY 984 -----TKAPKKTSTTKPKMTPRVKPKTTPTPKMTST--MPELN----- 1022
Db 1933 TIAAATAAP-----PRLPSP--PPEAVGSSTGVLPRLHMQMLQPPAAAV 1977
QY 1023 -----PTSRIABAMLQTTTRPNQTPNSKLVE----- 1048
Db 1978 AAPPPPPASSALVLQPPSPPPPPPPSQLLIQOASATYVSDMQDQLTYLVVAAMLITAVAA 2037
QY 1049 -----VNPKSEDAGGAEGETPHMLLRPHVFMPEVTPMDYLPBRVNPQGLIINP 1096
Db 2038 GRLLAAVLYRLLVSP-----PPHFLA-----PPRLE----- 2064
QY 1097 MLDSETNINCXKPVQGLTTLRNGTLVAFRGH----- 1127
Db 2065 -----TTIAGLILVALTYFSCMALGGPAADWHGSRTAAYCVLITIAVVP 2107
QY 1128 ---YFWMLSPFPSPPARRITEWGLP-----SPIDTVFTRCN----- 1162
Db 2108 YAAFLWLLA-----LARAMVPOFTLVEPMTTSSYASRPSAFERADTSPPRST 2157
QY 1163 -----CEGKTTFFPKDSOY-----WR-FT 1179
Db 2158 VSIATAADIGGTATGGGGAASGACADVCAAAAPATVAAGADVPSDDRKARGPHWKF- 2216
QY 1180 NDIKDAGVPKPIFKGFGGLTGOIQAALSTAKYKNWPESVYFFKRGSIQQYIYKPEPVOK 1239
Db 2217 ---DGVLPPTVASGVGGSGGVVPL-----PPLVA 2243
QY 1240 CPGRPALNYP-----VYGMTQVRRRRRFFAIGFSOTH 1273

```

| | | | |
|------------|--|--|---------------|
| Db | 2244 | LFGSGRHLPLPLPAAGATAAGGAATAPGSPROAAAGE | -----DTH 2287 |
| QY | 1274 | TIRIOVSRLAQVQKGLVHNEVKVLSILWR | -----1303 |
| Db | 2288 | -----QFGP-----H-----WKFSVPDGAATAGVGYDAAAGGAGGGGRG | 2324 |
| QY | 1304 | ----- | -----1303 |
| Db | 2325 | SRGSGTDGVRFGANGTRTPSDGAKGRSSHHGGGNGGNAMSSGAASFGASAGGGGAAAAASR | 2384 |
| QY | 1304 | -----GLP-----NVVTSALSL----- | -----1315 |
| Db | 2385 | LGSRSGSPGDGDYNAMSSGAAMRGGGAGASRFGPAGSRAGSPPEGAGGAGVAGGANAMFAGT | 2444 |
| QY | 1316 | -----DAPPPPLPAPGTVPPLPFANLLTSGADVRSRRALNAEVAAGG | 2504 |
| Db | 2445 | AGRAGALSPADGSSRRPGSGSNEGSRQMSGNAMTAGAVGVGGVRRPAGSLRNASD | 2504 |
| QY | 1323 | GVDYAFSKQDYNIDVPS-----RTARAI----- | -----1347 |
| Db | 2505 | G-----DAPPPPLPAPGTVPPLPFANLLTSGADVRSRRALNAEVAAGG | 2504 |
| QY | 1348 | -----TTRSG 1352 | ----- |
| Db | 2551 | GGGTSRSG 2559 | ----- |
| RESULT 15 | | | |
| SLP1_CLOTH | | | |
| ID | SLP1_CLOTH | STANDARD; | PRT; 1664 AA. |
| AC | Q06852; | | |
| DT | 01-JUN-1994 (Rel. 29, Created) | | |
| DT | 01-JUN-1994 (Rel. 29, Last sequence update) | | |
| DT | 05-JUL-2004 (Rel. 44, Last annotation update) | | |
| DE | Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1). | | |
| DE | protein 1). | | |
| GN | Name=olpB; | | |
| OS | Clostridium thermocellum. | | |
| OC | Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; | | |
| OC | Clostridium. | | |
| OX | NCBI_TaxID=1515; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=NCIB 10682; | | |
| RX | MEDLINE=9320931; PubMed=8458832; | | |
| PA | Fujino T., Beguin P., Aubert J.-P.; | | |
| RT | "Organization of a clostridium thermocellum gene cluster encoding the | | |
| RT | cellulosomal scaffolding protein Cipa and a protein possibly involved | | |
| RT | in attachment of the cellulosome to the cell surface." | | |
| RL | J. Bacteriol. 175:1891-1899(1993). | | |
| CC | -!- SUBUNIT: Assembled into mono-layered crystalline arrays. | | |
| CC | -!- SUBCELLULAR LOCATION: Cell wall. | | |
| CC | -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains. | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | |
| CC | use by non-profit institutions as long as its content is in no way | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | |
| CC | or send an email to license@ib-sib.ch). | | |
| CC | ----- | | |
| DR | EMBL; X67506; CAA47841.1; - | | |
| DR | PIR; T18262; T18262. | | |
| DR | InterPro; IPR008965; Cellul_bind. | | |
| DR | InterPro; IPR001119; SLH. | | |
| DR | Pfam; PF00395; SLH; 3. | | |
| DR | PROSITE; PS01072; SLH DOMAIN; 2. | | |
| KW | Cell wall; Repeat; S-layer; Signal. | | |
| FT | SIGNAL 1 28 Potential. | | |
| FT | CHAIN 29 1664 Cell surface glycoprotein 1. | | |
| FT | DOMAIN 36 763 4 X 156 AA approximate repeats. | | |
| FT | REPEAT 36 191 1. | | |

| | | | | |
|---|----------|--|------------|--|
| FT | REPEAT | 207 | 363 | 2. |
| FT | REPEAT | 409 | 565 | 3. |
| FT | REPEAT | 607 | 763 | 4. |
| FT | DOMAIN | 771 | 1377 | approximate tandem repeats of T-P-S-D-E- |
| FT | DOMAIN | | | P. |
| FT | DOMAIN | 1378 | 1449 | Gly/Pro/Ser/Thr-rich. |
| FT | DOMAIN | 1453 | 1494 | SLH 1 (incomplete). |
| FT | DOMAIN | 1495 | 1565 | SLH 2. |
| FT | DOMAIN | 1566 | 1625 | SLH 3. |
| FT | DOMAIN | 1626 | 1646 | SLH 4 (incomplete). |
| FT | SEQUENCE | 1664 | AA; 178194 | MW; 5F396695BA9FE74B CRC64; |
| Query Match | | | | |
| Best Local Similarity 16.4%; Score 1190.4; DB 1; Length 1664; | | | | |
| Matches 408; Conservative 145; Mismatches 465; Indels 946; Gaps 82; | | | | |
| QY | 4 | KTLPIYLLLLLVFV | ----- | -----LQVSSQE 26 |
| Db | 6 | KVLSILLTLLIITTSVNMSFAEATPSIEMVLDTXVHVDVITATIKVNNIRKLQAGY | 65 | |
| QY | 27 | LSCGRCFESFERGRCDCDAQCKYDKCPDYESFCAEVHNPTSPSSK | 82 | ---KAPPPS 82 |
| Db | 66 | LNIX | ----- | -----FDPEVLQVDPATGEEFTDKSMP 92 |
| QY | 83 | GASQTIKSTTKRSPKPNKTKKVIIESEITEHSVSENQESSSSSSSSSSSTIWKIK | 142 | |
| Db | 93 | VNRVLLTNSKYGTP | ----- | -----VAGNDIKSGIINFATGYNNLTAYK 131 |
| QY | 143 | SS | ----- | -----144 |
| Db | 132 | SSGIDHTGTIGRIGFKVLKKNKNTSIRFEDTSLMPGAIKNTSIFDMDAETITGVEVIQPD | 191 | |
| QY | 145 | -----KNSAANRELQK-KLKVKD | ----- | -----RTKKKT 172 |
| Db | 192 | LIVVEAPLADASVALELDTKVKVGDIIITATIKIENKKNFAGYQLNFKYDPTMLEAIEL | 251 | |
| QY | 173 | -----PKPPVVDAGSLDN | ----- | -----187 |
| Db | 252 | ETGSAIAKRTWPVTGGTVLQSDNYGKTTAVANDVGAGIINFADYNSLTLYRETGVABET | 311 | |
| QY | 188 | -----GDFKVTTPDT | ----- | -----STQHNK 205 |
| Db | 312 | GIIGKIGFRVLKAGSTAIRFEDTTAMPAGIEGTWFDWYGENIKGYSVVQGEIVAESEE | 371 | |
| QY | 206 | -----STSPKITTAKPINRPSLP | ----- | -----DTSK 231 |
| Db | 372 | PGEPTPEPVTETPVDPTFVTEEPVPSLPDYSVIMELDKTKVKVGDIIITATIKIENM | 431 | |
| QY | 232 | -----ETSLTVNKETTETKETTNTKQTSTDGKEKTS | ----- | -----265 |
| Db | 432 | KNFAGYQLNFKYDPTMLEAIELETSALAKRTWPTGGTV | ----- | -----LQSDNYGKTTAVAND 486 |
| QY | 266 | -----AKETQSIKTSKOLAPTQKVLAKPTPKAETTT | ----- | -----298 |
| Db | 487 | VGAGIINFADYNSLTLYRETGVABETGIIGKIGFRVLKAGSTAL | ----- | -----RFDTTAMPG 541 |
| QY | 299 | -----KGPAITTP | ----- | -----KEPTPT-TPKEPASTTPKEPTPT 328 |
| Db | 542 | AIEGTWFDWYGENIKGYSVVQGEIVAESEPTPEPVTETPVDPTFVTEEPVPSLP | 601 | |
| QY | 329 | ----- | ----- | -----328 |
| Db | 602 | DSYVIMELDKTKVKEGDVIIATIRVNNIKNLQAGYQIGIKYDKPKVLEAFNIETGDPIDEGT | 661 | |
| QY | 329 | ----- | ----- | -----T 329 |
| Db | 662 | WPAVGGTILKNRDLPTGVAINNVKSGIINFADYNSLTLYRETGVABETGIIGIIGRVR | 721 | |
| QY | 330 | IKSAPTTTK | ----- | -----EPATTTKSAPTTPKEPAPT 359 |
| Db | 722 | LKAEDTTIRFELESMPGSDIGTYMDVNLNRSIGVVIQAPAKAAS | 776 | -----DEPIPTD 776 |
| QY | 360 | TKEPAPTTPKEPA | ----- | -----PTTKEPAPTTPKSAPTTPKEPAPTTPKPAPTTPKEPAPT 413 |

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 27.7629 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-F
Perfect score: 72%
Sequence: 1 MAWKLPYLLLLSVFVIQ.....ARAITRSGQLTKVWYNCP 1363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.79.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 1348.5 | 18.5 | 3020 | 2 A43932 | mucin 2 precursor, |
| 2 | 1190.4 | 16.4 | 1664 | 2 T18262 | S-layer protein |
| 3 | 1065.3 | 14.6 | 2187 | 2 T30826 | nascent polypeptid |
| 4 | 1061.4 | 14.6 | 7962 | 2 I38346 | elastic titin - hu |
| 5 | 1052 | 14.5 | 3570 | 2 T45025 | mucin MUC5B, trach |
| 6 | 1023.2 | 14.1 | 1274 | 2 T16251 | hypothetical prote |
| 7 | 1020.2 | 14.0 | 1489 | 2 T31108 | cyst germination s |
| 8 | 983 | 13.5 | 1367 | 1 S48478 | glucan 1,4-alpha-g |
| 9 | 968.5 | 13.3 | 6642 | 2 T29757 | protein UNC-89 - C |
| 10 | 965.1 | 13.3 | 3507 | 2 T34513 | hypothetical prote |
| 11 | 919.5 | 12.6 | 5762 | 2 A41819 | proline-rich pepti |
| 12 | 898.5 | 12.3 | 1188 | 2 S49915 | extensin-like prot |
| 13 | 897.8 | 12.3 | 2897 | 2 B48666 | cell proliferation |
| 14 | 890.1 | 12.2 | 1229 | 2 T25697 | hypothetical prote |
| 15 | 885.9 | 12.2 | 1344 | 1 A35175 | mucin 1 precursor, |
| 16 | 885.5 | 12.2 | 3256 | 2 A48666 | cell proliferation |
| 17 | 879.3 | 12.1 | 4135 | 2 T42629 | tenascin-X - bovin |
| 18 | 861.3 | 11.8 | 1151 | 2 T18535 | high molecular mas |
| 19 | 851.5 | 11.7 | 3942 | 2 T42730 | Bassoon protein - |
| 20 | 849 | 11.7 | 2142 | 2 B35098 | MHC class III hist |
| 21 | 846.1 | 11.6 | 2232 | 2 T34434 | hypothetical prote |
| 22 | 838.3 | 11.5 | 5262 | 2 T03454 | mur protein - huma |
| 23 | 835 | 11.5 | 1832 | 2 T31113 | mucin-like glycopr |
| 24 | 831.3 | 11.4 | 3938 | 2 T42761 | Bassoon protein - |
| 25 | 826.5 | 11.4 | 1872 | 2 S36152 | MHC class III hist |
| 26 | 821.1 | 11.3 | 4548 | 1 S00657 | apoptosis(a) (EC |
| 27 | 817 | 11.2 | 4957 | 2 T03455 | ALR protein - huma |
| 28 | 815.8 | 11.2 | 5105 | 2 T32650 | hypothetical prote |
| 29 | 810.7 | 11.1 | 2774 | 2 A43359 | microtubule-associ |

ALIGNMENTS

RESULT 1

A43932

mucin 2 precursor, intestinal - human (fragments)

N:Alternate names: mucin SMUC-41

C:Species: Homo sapiens (man)

C>Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 09-Jul-2004

C:Accession: A49963; A45106; B45106; B33532; A61257; PQ0328; PQ0329

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.

J. Biol. Chem. 269, 2440-2446, 1994

A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the

A:Reference number: A49963; MUID:94132002; PMID:8300571

A:Accession: A49963

A:Molecule type: mRNA

A:Residues: 1-639 <GU1>

A:Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998

R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.

J. Biol. Chem. 267, 21375-21383, 1992

A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upst:

A:Reference number: A45106; MUID:93016075; PMID:1400449

A:Accession: A45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 626-1895 <GU2>

A:Cross-references: GB:M94131; NID:gi86395; PIDN:AAA59163.1; PID:gi86396

A>Note: sequence extracted from NCBI backbone (NCBIP:116706)

A:Accession: B45106

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 2037-3020 <GU3>

A:Cross-references: GB:M94132; NID:gi86397; PIDN:AAA59164.1; PID:gi86398

A:Experimental source: colon

A>Note: sequence extracted from NCBI backbone (NCBIP:116698)

R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.N.

J. Clin. Invest. 88, 1005-1013, 1991

A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymor:

A:Reference number: A43932; MUID:91358717; PMID:1885763

A:Accession: A43932

A:Molecule type: DNA

A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>

A:Cross-references: GB:M74027; NID:gi88863; PIDN:AAA59875.1; PID:gi88864

A>Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)

R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989

A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden

A:Reference number: A33532; MUID:89197956; PMID:2703501

A:Accession: B33532

A:Molecule type: mRNA

A:Residues: 1916-2193 <GU4>

A:Cross-references: GB:M22405; NID:gi88873; PIDN:AAA36334.1; PID:gi88874

A:Experimental source: intestine

R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

| | | | | |
|--|----|------|---|------|
| J. Clin. Invest. 87, 77-82, 1991 | Db | 1210 | SCVTNSSQVVCPEBEGKIINQTDGAFVCWBEICGNGTVEKHFNICSITRPSLTFTT | 1269 |
| A>Title: Human bronchus and intestine express the same mucin gene. | | | | |
| A'Reference number: A61257; MUID:91086481; PMID:1985113 | Qy | 242 | -----TVETKEITNTNKQISTDGKEKITSK- | 267 |
| A'Accession: A61257 | Db | 1270 | TITLPTTPTSFTTTTTTTTTSTVLSITTPKLCCLMSDWINEDHPSSGSDGDRPFDGV | 1329 |
| A>Status: not compared with conceptual translation | | | | |
| A'Molecule type: mRNA | Qy | 268 | --ETQSIETKSAD- | 279 |
| A'Residuals: 'T',1925-1948,'TTS',1952-1954 <JAN> | Db | 1330 | CGAPEDIERSVKDPHLSLEHQHQKQVQCVSVGFCIKNEDQFGNGFGLCYDKIRVNCC | 1389 |
| A'Experimental source: bronchus | | | | |
| R:Xu, G.; Huan, L.; Khatiri, I.; Sajjan, U.S.; McCoool, D.; Wang, D.; Jones, C.; Forstner, | Qy | 280 | -----LABTSKVLAKPTKAEITTKGPAITTPKBPPTTPPKBPASTTPKBPPTTKSA | 333 |
| Biocen. Biophys. Res. Commun. 183, 821-828, 1992 | Db | 1390 | WPMDKCIITPSPPTTTPSPPTTITLPLPTTTPSPPTTITTPPTTTPSPPTTTPPTTTP | 1449 |
| A'title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C- | Qy | 334 | P-TTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAP--TTP | 391 |
| A'Reference number: PQ0328; MUID:92198477; PMID:1550588 | Db | 1450 | PTTTPSPPTTTPPTTTPPTTTPSP--PTTTPSPPTTTPSPPT--TTTTTPPTTTPSPMTTP | 1507 |
| A'Accession: PQ0328 | Qy | 392 | -----KBPAPTTPKKAPATTTPKEPAPTTTPKEPTTPKEPAPTTTPKEPAPTTTPKEPAP | 443 |
| A'Molecule type: protein | Db | 1508 | ITPPASTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP | 1567 |
| A'Residuals: 2328-2342,'K',2344-2354 <XUG1> | Qy | 444 | TAPKAPATTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAAPTTPKEPAPTTTKS | 503 |
| A'Gene: GDB:MUC2 | Db | 1568 | TTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSP--PTT | 1624 |
| A'Cross-references: lrp15.5-11p15.5 | Qy | 504 | APTTPKEPSTTPKEPAPTTTPKEPAPTTTPKAP--TTPKEPAPTTTPKEPAPTTTPKAP | 562 |
| C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von | Db | 1625 | -PITP-----PTSTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT | 1679 |
| C:Keywords: glycoprotein; intestine; tandem repeat | Qy | 563 | APKEPAPTTTPKETAPTTPPKLTPTTPPEKLAPTTPKEPAPTTPELAPTTPPEEPTTTPPE | 622 |
| F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC> | Db | 1680 | TPSPPTTTP--SPPTT-----TMTTPS--PTTTPSSPITTTTTTPSSSTTP--SPPTTMT | 1730 |
| Query Match 18.5%; Score 1348.5; DB 2; Length 3020; | Qy | 623 | PAPTTPKAAAPNTPKPA-----PTTPKEPAPTTTPKEPAPTTTP-----KETAPTTP-- | 668 |
| Best Local Similarity 19.2%; Pred. No. 3e-20; | Db | 1731 | ESPTT-----TPSPPTTMTLPTTTPPTTTPSSPLTTTTPSPPTTTPPTTTPPTTTP | 1783 |
| Matches 479; Conservative 127; Mismatches 587; Indels 1305; Gaps 87; | Qy | 669 | -----KG----- | 670 |
| 649 WREHVCNKDVGCNSQVFLNLTTCQTCRSLSEADSHCLEGPAPVDGCGCPDHTFLDE | Db | 1784 | PLCNWTGLDSGKPNFHKPGDGTDELIGDVGCPGWAANISCRATMYDPVPIQLQVTVCD | 1843 |
| 3 WK-----TLPIYILLLLSVFVIQVSSQELSCKRCFESF-----E 38 | Qy | 671 | -----TAPTTLKBPAPTTPKKAPKELAP | 694 |
| 649 WREHVCNKDVGCNSQVFLNLTTCQTCRSLSEADSHCLEGPAPVDGCGCPDHTFLDE | Db | 1844 | VSVGLICKNEQKPGVLPMAFCLNVEINVQCECVQPTM--TTITTENPTPTTTP | 1900 |
| 39 RGR-----ECDC-----DAQKYDKC-CPDYESFCAEVH-----67 | Qy | 695 | TT--TKBPTSTSDKP-----APT--TPKGTAPTTPKBPAPTTPKBPAPTTPKGTAPTTLKE | 747 |
| 709 KGRCVPLAKSCYHRGYLEAGDVVVRQBERCVRDGLRHLRQLRLIQSGCTAPKIHWC | Db | 1901 | TTTTPPTPTSTQSPNGLOAPTPTPTTSTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTP | 1959 |
| 68 -NTPSPSSKAPPPGASQTIKS-----TKRSPKPPN 100 | Qy | 748 | PAPTTPKPAKELAPTTPKGTST--TSDKAPAPTTPKETAPTTPKEPAPTTPKBP-- | 801 |
| 769 SNLTALATSK---PRAISQTLAAGYHTECVGCVCPDGLMDDRGCGVVEKECP---821 | Db | 1960 | PTPTGTPTTLVTTTTTTTTTMTPTPTPTSTKSTTVPTTITTTTATPTPTGTQTMPI | 2019 |
| 101 KKTKKVISEITEHSVSNQSSSSSSSSSSSSSTTWIKSKNSAA-----149 | Qy | 802 | --APTTPPTTPPTTSEVSTPTTK-----BPTTIHKSPDESPELSABPTPKALNSPKE | 854 |
| 822 -----CVHNDLYSSGA-----KIKVDCNTCTCKRGRWVCTQA 854 | Db | 2020 | STTTTTPPTTPTTGTSTGPTTHTSTAPIAELTTSNPPDESSTPQTSRSTSSPLTSTTLL | 2079 |
| 150 -----NREL-----153 | Qy | 855 | PGVPTTKTPAATKBPMTTAKDKITERDLRTTPTETTTAAPKMTKETATTEKTESKITA | 914 |
| 855 VHGTCISYSGHYITFDGKYDFDGHGCSYVAVQYCGQNSLSGFSIITENVPCGTGV 914 | Db | 2080 | STLPPAIEMTSTAPSTPTAPT--SGGHTLSPPTTTPPTTTPPTTTPPTTTPPTTTPPTT | 2137 |
| 154 -----QKXIKVXDNKK-----NRT 167 | Qy | 915 | TTTQVSTTTTQDTPPKITT--LKTITTLAPKVTITTKTITITTEIMNKPEE-----962 | |
| 915 TCSKAIKIFMGRTLEKLEDKHRVVIQRDEGHVAVTTREVQYLVVSSSTGLIIVWDKRT 974 | Db | 2138 | STVQTTTISAMTPTPTPLSTPSIIRTTGLRYPVSSVLCCVLNDIYVYAPGEVYNGTYGD | 2197 |
| 168 -----KKKPTPKPPVDEAGSL-----DNGDFKV-----TTPD 196 | Qy | 963 | -----TAKPKDRATNSKATTPKQKPTKAPKXKSTKPKXTMTRV 1002 | |
| 975 TVFIKLAPSYKGTV-----CGLCGNFDRHSNNDFTTRDHMVVSSBELDFGNSWKEAPTCPD 1029 | Db | 2198 | TCYFVNCSLSTLEFYNWSCFSTPTPTPSK--STPTPSKPSSTPSKPTPGTKPPECDF | 2256 |
| 197 TSTQT-----HNKYSTP-----209 | Qy | 1003 | RPKTTPTPRKMSTSNPELNTSRIAEAMLOTTTTRNOTPNSKLVENP-----1051 | |
| 1030 VSTNPEPCSLNPHRRSWAEKQCSILKSSVFSICHKVPKPFYEAACHVDSKSCDGGDCE 1089 | Db | | | |
| 210 -----KITAKPIN 218 | Qy | | | |
| 1090 CFCSAVAYAQCTKEGACVFWRTPDLCPIFCDYNNPHECEWHVEPCGNRSFETCRTIN 1149 | Db | | | |
| 219 -----PRPSLP-----224 | Qy | | | |
| 1150 GIHSNISVYLEGCVPRCPKDRPIVEEDLKCVTADKCGCYVEDHTYPPGASVPIETCK 1209 | Db | | | |
| 225 -----PNSDTSKE---TSLTVNKET-----241 | Qy | | | |

```

Db 2257 DDDR-----QENETWLCDFM-ATCKYNNNTVEIVKVECEPPMPETCSNGLQ 2302
QY 1052 ---KSDAGGA-----EGETPHMLLRPHVMPVPTDM 1081
Db 2303 PVRVEDPGCWHWCDCYCTGWDGPHVYTFGLYISYQGNCTYVIVE-----EISPSV 2356
QY 1082 DYLRVPVNOGIIINPMLSDETNICNGKPVDTGLTLRN-----1118
Db 2357 D-----NFGVYIDNYHCD-----PNDKVSQCRTLIVRHETQEVLLIKTVHVMQVQ 2402
QY 1119 -----GTLVAFRG-----HYF-----1129
Db 2403 VQVNRQAVLPYKXYGLEVYQSGINVVVDIIPELGLVSYNGLSFSVRLPYHRFGNNTKGQ 2462
QY 1130 -----WMLSPESP-----1138
Db 2463 CGTCNTTSDCILPSGEIVSNCERAAQWLVDPSKPHCHSHSSTTKRPVTVPGGKT 2522
QY 1139 SPARRITEVMGIPSI-----DTVTRCN-----CEGKTFPFKDSQY-----1175
Db 2523 TPHKDCST-----PSPLCOLIKDSLFAQCHALVPPQHYVDACVDFSCFMPGSSLECASLQA 2577
QY 1176 -----WR-----1177
Db 2578 YAALCAQONICLDNRNHTHGACLVCEPSSHREYQACPAEPEPTCKSSSQNNNTVLVEGCF 2637
QY 1178 -----FTNDIKDAGY--PKPIPKFGGLTGQIVAAALSTAKYKNWPESVYFFK--1222
Db 2638 CPEGTMYAFGFDVCVKTGCGVDNVPREG-----EHFEFD 2675
QY 1223 -----RGS-----IQYIYKQEPVQK-----1240
Db 2676 CKNCVLEGSGIICQPKRCSQKPVTHCVBDGYLATEVNPADCCNITVCKNCTSLCKE 2735
QY 1241 -----PGR-----RPALNPYVYEMTQ-----1257
Db 2736 KPSVCPILGFEVKSQWVPGRCPPFYWCESKGVCHGNAEYQP--GSPVYSKQCCVCCTDK 2793
QY 1258 -----VRRRRFERAIGPSQTH 1273
Db 2794 VDNNTLLNVIACTHVPCTSCSPGFELMEAPGECCKCEQTHCIIR-----PDNQH 2845
QY 1274 TIRIQSPARLAYODKG-----VLHNEVKYSILWRGLNPVVTSAISLPN-----IRKP 1321
Db 2846 VI---LKPQDFKSDPKNCTFFSCVKIHNQLISSV-----SNITCFNDFASICIP 2892
QY 1322 DGYDYVAFSKQDYINIDVPSRTARATTR-----S 1351
Db 2893 GSITF-----MPNGCCKTCFPRNETRVPVPCSTVPVTVTSYAGCTKTVMNHCS 2940
QY 1352 GQTLKSVWY-----NCP 1363
Db 2941 GSCGTFVWYSAKAQAALDHSCCKEETKSQREVVLSCP 2978

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; PMID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUC>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841
Query Match 16.4%; Score 1190.4; DB 2; Length 1664;

```

```

Best Local Similarity 20.8%; Pred. No. 2e-17;
Matches 408; Conservative 145; Mismatches 465; Indels 946; Gaps 82;
QY 4 KTLPIVLLLLLSVFV-----10QVSSOE 26
Db 6 KVLISLLTLLLLISITSVNMSFAEATPSIEMVLDKTEVHVGDVITATIKVNNIRKLAGVQ 65
QY 27 LSCKGRCPESPERGRECDCAQCKKYDKCCPDYSEFCAEVHNPTSPSSK---KAPPS 82
Db 66 LNIK-----FDEVLQPDVPAIGSEFTDKSMP--92
QY 83 GASQTIKSTTKRSPKPNKKTKKVIESEBITEHHSVENQESSSSSSSSSTTIWKIK 142
Db 93 -VNRVLLTNSKYGTP-----VAGNDIKSGIINFATGYNLTAYK 131
QY 143 SS-----PKPPVVDGAGSLDN-----144
Db 132 SSGIDEHTGIIIGEIGFKVLKKQNTSIRPDTLSMPGALSGLDFMDAETITGYEVIQPD 191
QY 145 -----KNSAANRELQK-KLKVKD-----NKN-----RTKKKPT--172
Db 192 LIVVEAPLKDASVALELDKTKVKGDIITATIKIENNMKNFAGYQLNKKYDPTMLEALEL 251
QY 173 -----PKPPVVDGAGSLDN-----187
Db 252 ETGSAIAKRTWVTVGTGTVLQSDNYGKTTAVANDVGAGIINFAPAEAYSNLTKYRETVABET 311
QY 188 -----GDFKVTTPDT-----STQHNKV-----205
Db 312 GLIGKIGFRVLKAGSTAIRPEDITTAMPGAIECTWFDWYGENIKGYSVVQGEIVAESEE 371
QY 206 ---STSPKITTAKPINRPSLP-----PNS-----DTSK-----231
Db 372 PGEETPEEPVPTPVDPTVTVEEPVSELPDSYVIMELDKTKVKVGDIIITAIKINM 431
QY 232 -----ETSLTVNKETIVETKETTITNTKQTDGKEKTS-----265
Db 432 KNFAGYQLNKYDPTMLEALETGSIAKRTWPTGTV-----LQSDNYGKTTAVAND 486
QY 266 -----AKETQSIEKTSAKDLAPTSKVLAKPTKAEITT-----298
Db 487 VGAGIINFAPAEAYSNLTKYRETVABETGLIGKIGFRVLKAGSTAI-----RFEETAMPG 541
QY 299 -----KGPALTTP-----KEPTPT-TPKEPASTTPKEPTPT--328
Db 542 AIBGTYMFDWYGENIKGYSVVQGEIVAEGETPEEPVPTPTVPTPTTTEPVPSEL 601
QY 329 -----328
Db 602 DSYVIMELDKTKVKEGDIITIRVNNIKNLAGYQIGKYDPKVLAFNIEGDPIDEGT 661
QY 329 -----T 329
Db 662 WPAVGOTILKNRDYLTGVAINNVSKGILNPAAYVYVFDYREEGKSEDTGIIGNIGFRV 721
QY 330 IKSAPTPK-----EPATTTKSAPTTPEKAPT 359
Db 722 LKAEDTTIRFEELSMPSGIDGTYMLDYNLRISGVYVWIQAPAKAAS-----DEPIPTD 776
QY 360 TKEPAPTTPKEA---PTTKAPATTTKSAPTTPEKAPTTPKAPATTTPKEAPT--413
Db 777 TFSDEPTPSDEPTPSDEPTPSDEPTSETPEEPITDTPSDEPTPSDEPTPSDE 836
QY 414 -----TPK-BPTPT-TPKEAPT--TKEPAPTTPKEAPTAPKKA--PTTPKEAPT-460
Db 837 PTPSDEPTPSDEPTPSDEPTPEEIPDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 896
QY 461 TPKEPAPTTPKPSPTTPKEPAPTTPKAPATTTKSAPTTTKAPATTTKSAPTTKEPSPTTKEDA 520
Db 897 TPEEPIPTDTPSDEPTPSDEPTPS--DEPTPSDEPTPSDEPTPSDEPTPEEIPDTPSDE 953
QY 521 PTTTPKEAPTTPKKA-----PTTPKEAPT--TPKEPAPTTPKKAAPTAPK 565

```



```
QY 337 PKK-PAPTTTKSAPTPK-----EPAPTTTKEPAPTTTKEPAPTTTKEPAP----- 381
Db 980 PKKSPKAASKTATPSPGVTAVPLEIPPCSKKAPKTAAPKESATSSSRAPKTAVS 1039
QY 382 -----TTTKSAPTTKEPAPTTPKK-PAPTTTPKE----- 409
Db 1040 KEIPSGVTAVPLEISLPLKETSKA--TPGEKSASSPKRSPKTAGPKETPPGGVTAVPP 1097
QY 410 ----PAPTTTPKEPT-----TTPKE-----PAPTTTKE 432
Db 1098 EISLPPKETPQWATPNESLAASSQKRSKTSVPKETPPGGVTAVPLEIPSPQAKPKTAV 1157
QY 433 P-----AFTTPKEPAPTAPKAPATT----- 454
Db 1158 PKQIPTPBDVAITLAGSPLSPKASKTAAPKEAPATPSGVTAIVSGEISPSPKTSKTA 1217
QY 455 -----KEPAPTTTKEPAPTTTKE-----PSPTTTPKE-----PAPTTTTSAP 490
Db 1218 PKENSATLPPKRSKTAAPKE--TATSSGVTAVFSEISPSPTPASKGVPTLTLPKGAP 1276
QY 491 TTTKEPAPTTTTSAPTT--PKEPS--PTTTKEPAPTTTKEPAPTTTKEPAPTT----- 539
Db 1277 NALAE--SPASPKVPKTAAPETSTPSQKIPKVGAPKEASATPPSKTPTKTAVPKETS 1335
QY 540 -PKEPAPTTTKEPAPTTTKEPAPTAPE--PAPTTPKETAPTTPKKLTPTTPEKLAPTTPE 597
Db 1336 APSEGVTAVPLEIPSPKAPKTAAPKETPAPS--PEGATTAPVQI--PPSPKSKKAGS 1392
QY 598 KPAPTTPEELAPTTPEPTTTPREP--APTTPKAAP-----NTP 636
Db 1393 KETPTTP-----SPEGVTAAPLEIPISSKTSKVASPKETLVTTPSSKSLSQTVGPKETS 1446
QY 637 KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPKELAP-- 694
Db 1447 LEGATAVPLEIPSHKAPKTVDPKQVPLTPSPKDAPTTLAE--SPSPSKK--APKTAAPPS 1504
QY 695 --TTTKEPTTSGDKAPTTPKGAPTTTKEPAPTTTKEPAPTTTKEPAPTT-----PTTPKGAPT 743
Db 1505 ERVTTVTP-----EKPA--TPQKASGTTASKVPVPAETQEVAVSSRETPTVPFAPVPVKNS 1558
QY 744 TLKEPAPTTTKEPAPKELATTTKGTSTTSKAPATPTTKEPAPTTTKEPAPTTTKEPAPTT 803
Db 1559 SHAKTSKTIELKAPATLPSPTKSPKIPSSKK-----APRTSAP-----KEFP 1602
QY 804 TTPETPTTSEVSTPTTKEPTTIHKSPPDESPE--LSAEPKPALENSPKPGVPTTK 861
Db 1603 ASPSIKPVITSLAQT-----APSLQKAPSTTIPKENLAA--PAVLPSKSKSPAAP-AR 1653
QY 862 TPAATPEMTTAKOKTTERDLRTP-----ETTTA-----APKMTKETATTT 904
Db 1654 ASASLSP-----ATAAPQATAPKEATTIPSCKAAATETPIETSTAPSLGAPKETSETSV-- 1708
QY 905 EKTTESKITATTTQVSTTQDTTPKLTIL-----KTTTLAPKVTKTTKTTTTEIMNKP 960
Db 1709 -----SKVLMSPGPKKASSKSRASTLPATTLPSLKEASVLSPTATSSGK----- 1752
QY 961 BETAKPKDRATNSKATTPKQPKTAPKK----- 989
Db 1753 DSHISPVSDACSTGTT--PQASEKLPKSGKPTAFTEMLAAPAPESALAITAPIQSPGA 1810
QY 990 -----PTSTKPKTMR--VRKPKTPTTPKMTSTMTPELNPPTSRIAEAMLQTT 1035
Db 1811 NSNSASSKCPDPSSKDKTKGLPSAVALAPQTVPEK-----DTSKAETILLVSP 1860
QY 1036 TRPNQ-----TPNSKILV 1047
Db 1861 AKGSDCLHSPKGVSGQVATPIAAFTSKVPPEAVSASVAPKAPAPASLTLAPSPVAPLP 1920
QY 1048 EYNPKSDEAGGAETGPHMLLRPHVPMPEVTPDMDYLPVFPVQGIILNPLWDEINIENG 1107
Db 1921 PKQPLLESAPGSLSPSKL-----PVPABEDEL-----PLIPPEA--VSGG 1961
QY 1108 KPVGDGLTLRLNGTLVAFRGHYFWMILSPFSF-----PSFARRITEVMGIPSIDTVTR 1160
```

```
Db 1962 E-----PQPIILNNPAPKPA-----GTPAPAPGA----- 1986
QY 1161 CNCEGKTTFFKDSQYWRFTNDIKDAGYKPKPIFKGG----- 1197
Db 1987 -----KQPVILKNNKSGTSDSDSDESPLEEQDSTQTA 2019
QY 1198 -LTGQIVAAALSTAKYKNWPESVYFFKRGSGSQYIYKQBPVKQKCGRRPALNYPVVGEMT 1256
Db 2020 TQQAQAAAAE-----IDEEPVSKAQSR-----S 2044
QY 1257 QVRRRRFERAIGPSQTH-----TIRIQVSPARLAYQDKGVILHNEVKVSILMRGLEPNVVS 1311
Db 2045 EKKARKAMSKLGLQVGTGVTIR-----KSKNILF-----VIT- 2079
QY 1312 AISLPIRKPKDGY-----DYAFSKQOYVNIQVPSRTARAITTR--SQITLSKVVYN 1361
Db 2080 -----KPDVYKSPASDITVIFGEAKIEDLSQQAQAAAAEKPKVQGEAVSNIQEN 2128

RESULT 4
138346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C:Accession: I38346
R:Labett, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38346
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g10
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:I127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 14.6%; Score 1061.4; DB 2; Length 7962;
Best Local Similarity 15.3%; Pred. No. 3.3e-13;
Matches 429; Conservative 152; Mismatches 508; Indels 1716; Gaps 106;

QY 2 AWK-----TLPIYLLLLSVFVIOQVSSQ----- 25
Db 5411 AWEBHLQDVTLKEGQCTMTVQ-----FSPVNVKSEWFRNGRIILKQGRHKTVEHKV 5463
QY 26 -----ELSCGRCFESPERGECDCQCKKYDK-----CCPDYESFCAE- 65
Db 5464 HKLTIADVRAEDQGGYTCK--YEDLETSALRLIEAEPIQFTKRIQNIIVVSEHSQATFEC 5520
QY 66 -----VHN----- 68
Db 5521 EVSPDDAIVTYKGPTELTSOKYNFRNDGRCHYMTIHNVTDPDEGVSVIARLEPRGEA 5580
QY 69 -----PTSP-----PSSK 76
Db 5581 RSTAEIYLTKEIKLELPPDIPDSRVPIPTWPIRAVPPEEIPPVVAPVPLLPTPEEK 5640
QY 77 KAPP-----PSGASQTI----- 88
Db 5641 KPPPKRIEVTKAVKADAKVAVKPKEMTPREEIVKKPPPTTLIPAKAPEIIDVSSKAE 5700
QY 89 ----- 88
Db 5701 EVKIMTIITRKKEVQKEAVVEKKQAVHKEKRVFIESFPEPYDELEVPYTFEPFQPYE 5760
QY 89 ----- 88
Db 5761 EPDEYEEIKVEAKVEHVEEEDFEEQYEREEGYDEGEHEEAEYQREVIOQKE 5820
QY 89 -----KSTTKRSPKPPN--KK-----KTKKVISEBEI----- 113
```

```
Db      5821 VVEESHKVPKAPVPEKAPPPKVIKKPVIEKIEKTSRMBEEKVQVTKVPEVSKKIIVP 5880
QY      114 -----T-EHSVSENQESSSSSSSSSSSTIWKI----- 141
Db      5881 QKPSRTPVQEEVIEVUPAVHTKKWVISBEKMFASHTEEVSVTVPEVQKEIVTTEKIH 5940
QY      142 -----KSSKNSAANRELQKKLVKDNKNRTKKP----- 171
Db      5941 VAVSKRVPEPPKVPPELPEKPAPEVAPVPIPKKVEPPAPKVPPEVPPKVPPEEKPPVPVK 6000
QY      172 -----TPKPPVDEAGSLDNGDFKVTTPDTSIQ-----HNKVESTPKI 211
Db      6001 KEPAAPPKVPEVPKVPES-----KIPVPVAKKEAPPAKVPVEQKGVVTEEKI 6050
QY      212 TTA-----KPINPR-----PSL----- 223
Db      6051 TIIVQREESPPPAVPEIPKKKVPPEEKPV-PRKEEEVPPPPKVPALPKKVPPEEKVAVPV 6109
QY      224 -----PPNSDTSKET----- 233
Db      6110 PVAKAPPRAVSKTWEKRFVABEKLFAVQRFVETRHEVSAEEMSYSSEBEGV 6169
QY      234 SLTV-NKETTVETKETTINKQSTDCKEKTSAK-----EQSIEKT----- 275
Db      6170 SISVYREEREEREAEVTEYVMEEPYVVEEKLHIISKRVAEAEVTERQEKKIVL 6229
QY      276 -----SAKDIAPTSKVLAKPPPKAEITTKGPALATPKPEPTTPPKP 317
Db      6230 KPKIPAKIEBPPPAKVPEAKKVIPEKKVPA-PVPKKE-----KVPPKVPPEP 6277
QY      318 ASITP-----KEPTP-----TTIKSAPTTKBPATTTKS 347
Db      6278 KKPVEKKVPKVKIMBEELPAKVTEKHMQITOEEKVLVAVTKEAP-PRARVPEEPKR 6335
QY      348 A-----PTTPKBPATTT-----KEBAP----- 365
Db      6336 AVPEEKVLKLPKREBPPPAKVTEFRKRVKVEKSVSEAPKREPOIKEVTIMEEKERAY 6395
QY      366 -----TPPK 369
Db      6396 TLEEAVSQREEEYEEYDYKEFBEYBTEEDQYEEYEREYEREYEEHVEITEPE 6455
QY      370 EPAPT--TTKEBAPTITKSAP-----TTPKE--PAPTPKBPATTTKEP----- 410
Db      6456 KPIPVKVPPEBVPETKAPDAKVLKKAVPBEKVVPPIPKLKPVPKVPPEEPKKVPEEK 6515
QY      411 -----APTTPKE----- 417
Db      6516 IHSITKREKEQVTEPAKVPKRVVABEKKVPVPRKEVAPPVPRVPEVPEKELEPEEVAF 6575
QY      418 -----PTPTPKBPATTTKEBAPTTPK 439
Db      6576 EEEVTVHEEYIEEYEEYIHEEEBFITEEYVVPVPIPVKVEVPRKVPPEEKPPVPVKK 6635
QY      440 BPAPTA-----PKPK-----APTTPKEBAPTTPKBPATTTKEPS-- 474
Db      6636 KEAPPKAPVPEVPKPEEKVPVPLIPKKEKPPPAKVPEVPEKPPV-BEKVPVPVPKVEAPP 6694
QY      475 ---PTTPKEBAPTITTKSAPTITTKBPATTTKSAPTTPKE-----PSPTTTKEBAPT 522
Db      6695 AKVPEVPKPPVPEKKVPVPAKKEVAPPA--KVPEVPKLLIPEEKKPTFPVKKEVAPPK 6752
QY      523 TPK--EPAP----- 529
Db      6753 VPKEEPVPVVALQEBEVLPEEBIVPEEVLPEEBEVLPEEBEVLPEEBEVLPEEBEVLPEEBEVL 6812
QY      530 -----TTPKKBAPT----- 538
Db      6813 PPEEEVPPPEEYVPEEEFVPEEVLPEVKPKVPVAPVPEIKKKVTEKKVVIKKEEA 6872
QY      539 -----TPKE-----PAPT-----KPEP----- 551
```

QY 1235 EPVQKCGRRPALNYPVYGMETQVRRRRF-----EKAIGPSQTHIRIQSPA----- 1282
 Db 7848 SP-----KHFIADGDKRKU-----HIIDVQLSDAGEYTCVL 7879
 QY 1283 RLAYQDK-----GVLHNEVKY-----SILWRGLPNVVT 1310
 Db 7880 RLGNKEKTSTAKLVVEELPFVKLTBEVTVWKGQPLYLSCELNKRVDVWR----- 7932
 QY 1311 SAISLPLNRKPDGYDYAFSXDQYINIDVPSRTA-----RAIT 1348
 Db 7933 -----KDGKIVVEKPGRIVPVGVIGLMRALT 7957

RESULT 5
 T45025
 mucin MUC5B, tracheobronchial [imported] - human (fragment)
 C;Species: Homo sapiens (man)
 C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C;Accession: T45025
 J;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 J. Biol. Chem. 272, 3168-3178, 1997
 A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
 A;Reference number: Z22899; MUID:97166151; PMID:9013550
 A;Accession: T45025
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-3570 <DES>
 A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
 A;Experimental source: placenta
 C;Genetics:
 A;Gene: MUC5B

Query Match 14.5%; Score 1052; DB 2; Length 3570;
 Best Local Similarity 13.3%; Pred. No. 9.5e-14;
 Matches 492; Conservative 129; Mismatches 536; Indels 2550; Gaps 110;

QY 6 LPYLLALLSVFVIQVSSQELSCK-----GRCFESPER----- 39
 Db 3 LPV-----STVCVREV-----CRSSWYNGHRPEPGLGGGFETENLRQGYQCPV 50
 QY 40 -----GRECDCAQ-----CKKYDKCCP 57
 Db 51 LADIECRAAGLPDMPLEELGQVDCDRMGLMCANSSQSPPLCHDYELVLCCEYVPCGP 110
 QY 58 ----- 57
 Db 111 SPAGTSPQPSLSASTEPAVPTPTQTATEKTTLWVTPSIRSTAALTSQTGSSSGFVVT 170
 QY 58 -----DY-----CAEVH----- 67
 Db 171 PSAPGTTTCQPRCQWTEFDEYPSKSEQLGDVSYDKIRAAAGHLCQPKDIECQAESF 230
 QY 63 ----- 67
 Db 231 PNWTLAQVGQKVCHDVHGLVCRNWEQGVFKMCMYRIVLCCSDDHCKGRATTPPT 290
 QY 68 -----NPTSPSSKKAPPS-----GA 84
 Db 291 ELETATTTTQALFSTPQTSFGLTRAPPASTAVPTLSEGLTSPTYTSLTGATGGP 350
 QY 85 SQTIKST-----TKRSPKPPNKKTKKVE----- 109
 Db 351 RQSAGSTEPTVGAVATSLPTRSALPGTIGSLGTWEPSPQPTLAPTMATSRAPRTGAS 410
 QY 110 --SEE-----ITEHVSVENQES----- 125
 Db 411 TASKEPLTSLAPTLTSLTSQAETSPTRTETMTSPLNTTTSQGTTRCQPKCEWTF 470
 QY 126 ----- 125
 Db 471 DVDFPTSGVAGDMETFENIRAAAGKMCWAPKSIIECRAENYPEVSIQVGVLTCSLETG 530

QY 126 -----SSSSSSSSSSSTIWKIKSKNSA 148
 Db 531 LTKNEDQTRFNMCFNYNRVVLCDDYSHCPSTILATSSATPSSTPGCTTWILT----- 584
 QY 149 ANRELQKKLKVCKNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVT--TPDTSITICHNKVS 206
 Db 585 -----KPTTATTASTGSTATASSTQATAGPHVSTT-----A 618
 QY 207 TSPKITTAK-----PINRPSLPNDSOTSKETSL----- 235
 Db 619 TTPVTSSKATPFSSPGTATAPALRSTATTPTATSTAIPTSSSLGTTWRLSQTTPMA 678
 QY 236 ---TVNKETIVETKET---TTNKQISTDKKXKTSKAKETQSIKTSKADLAPTSKVL- 287
 Db 679 TMSATPSSTPETVHTSTVLTATTATTGATGAVATPS-----STPGTAHTTKVLT 728
 QY 288 -----AKP----- 290
 Db 729 TTTTGTATSSSPGRARTLPVWISTTTTPTRGRTVTPSSIPGTHHTPTVLTATTTTVA 788
 QY 291 -----TPKAEITTKG--PALTTTKEPTPTT-----PKEPASTTTPKEPTTTIKSAP 334
 Db 789 TGSMAFSSSTQISGTTPSLTT---TATTITATGSTTNPSSTPGITPIPVLTATTATP 844
 QY 335 -----TPKEPAPTTTKS-----APTTP-----KE 354
 Db 845 AATSSVTTPSSALGTTHTTPFPVNTTATTHGRSLSPSPSHVTVCTAWTSATSGILGTHITE 904
 QY 355 PAPTTTKEPAPTT-----PKEPAPTTTKEPAPTTTKSAPT----- 390
 Db 905 PSTGSHTPAATTGTQHSTPALSSHPSSRTTESPPSGTTTGTGHTTATSRATATPS 964
 QY 391 -----PKEP--APT----- 398
 Db 965 KTRTSTLLPQOPTSAPITTVTMGCEPQCAWSEWLDYSYPMWPGSGDFDTYSNIRAAGG 1024
 QY 399 -----PKKP----- 402
 Db 1025 AVCEQPLGLECRAOAGFVPLRELGQVVEGSLDFGLVCRNREQVKFKMCFNVEIRVFCC 1084
 QY 403 ---APTTPKEPAPTTTPE-----PTTPK-----PTTPTPKEPAP--T 429
 Db 1085 NYGHCSTTATSTATPSSTPGTTWILTATTATTESTGSTATPTSTLRTATPPPKVLT 1144
 QY 430 TKEPAPT-----TPKEPAPTAPKDA---PTTPK-----EPAP-----TT 461
 Db 1145 TTATTTVTSSKATPSSPGTATAPALRSTATTPTATSVTPIPSSSLGTTWRLSQTIT 1204
 QY 462 PKEPAPTTTKEPSP-----TPKEPAPTTTKSAPTTTKEPAPTTT-- 501
 Db 1205 PTATMSTATPSSTPETAHTSTVLTATATTGATGSAVATPSSTPGTAHTTKVPTTTTGT 1264
 QY 502 ---KSAPTTPKEP-----SPTT----- 515
 Db 1265 ATPSSSPGTLTPPWISTTTPTTRGRTVTPSSIPGTHHTATVLTATTTTVAIGSMATP 1324
 QY 516 -----TKEPAPTTTKEPAP-----TPPKAPPT-----TPKE 542
 Db 1325 SSSTQTSQTPPSLTTATTITATGSTTNPSSTPGIRPIPVLTATTATTAASSTVTPSS 1384
 QY 543 PAPTTPKEPAPTT----- 556
 Db 1385 ALGTHTPPEVNTTATTHGRSLSPSPSHVTVTAWTSATSGILGTHITEPSTGTSHTPAA 1444
 QY 557 -----KXKAPT-----APKEPAPTT-----KETAPTTPK-----ITPTT 587
 Db 1445 TGTGTTQHSTPALSSHPSSRTTESPPSGCTTTPGHTTATSRATATATPKTRTSLTSS 1504
 QY 588 PEKLAPTTP-----EXP----- 599
 Db 1505 PTS-APITTVTMGCEPQCAWSEWLDYSYPMWPGSGDFDTYSNIRAAGVACEPLGLE 1563
 QY 600 -----APTTP-----APTTP 605

```
Db      1564 CRAQAQPGVPLRELQGVVECSLDGLVLCRNREQVGKFKMCFNVEIRVFCNCYGHCESTPA 1623
      QY      606 ELAPTTPEE-----PTPTTPEEPAP-----TTP----- 628
Db      1624 TSSATSPSPGTTWLTTEQTTAAITATGSAIPSSPGTAPPPKVLTSQATITFATS 1683
      QY      629 -KAAAPNTPK-----EPAPT-----PKEP----- 647
Db      1684 SKATSSSPRTATTLPLVLTSTATKSTATSTPTIPSSTLGTGTGTSQNRPPHPMATMSTIHP 1743
      QY      648 -----APTTPKEPAPTTPKETAPT-----PKGTAPT 675
Db      1744 SSTPETHTSVLTTKATITTRATSSMSSTSPSTGCTTWILTELTAATTAALPHGIPSS 1803
      QY      676 -----LKEPAPT-----TPPK----- 686
Db      1804 PGTTWILTEBTTATVPTPGSTATASSRATAGTLKVLTTATPTVISSRATPSSPG 1863
      QY      687 -----PAKELAPT----- 695
Db      1864 TATALPALRSTATTPTATSVTAIPSSSLGTAWRLSQTTTTPTATMSTATSPETVHTS 1923
      QY      696 -----TTKEPTSTT-----SDKP 708
Db      1924 TVLTATTATTRGSAVPSPSTPGTAHTTKVPTTTTGTATPSSPGTALTTPVWISTTT 1983
      QY      709 APTT----- 712
Db      1984 TPTIRGSTVPSSIPGTHTHATVLTITTTIVATGSMATPSSSTQTSCTPPSLATTATTIT 2043
      QY      713 -----PKGTAPTTPKEPAPT----- 728
Db      2044 ATGSTINPSSPGTTPPPVLTITATTTPAATSVTPSSALGTHTPPVENVTTATHGRS 2103
      QY      729 --PKEP--APT-----TPKGTAPTT-----LKEPAP-----T 751
Db      2104 LPSSPHVTPAWTASGILGTHITHTEPSTGISHTPAAATGTTQPSPALSSHPSSRT 2163
      QY      752 TPKPAPAKELAPTTTKGTSTSDKPAPTTPKET-----AP-----TPKPEP- 793
Db      2164 TESPPSP-----GTTTPGHTRGTSRTTATATPASKURTSTLLPSSPTSAPITTVVTTGCEPQ 2219
      QY      794 ----- 793
Db      2220 CAWSEWLDYSYMPGPGGGDFDYSNIRAAGAVCEQPLGLECRAQAQPGVPLRELQGV 2279
      QY      794 -----APTTPKPAKAPTTPETPPPT----- 813
Db      2280 ECSLDGLVLCRNREQVGKFKMCFNVEIRVFCNCYGHCESTPATSTATPSTPGTTWILT 2339
      QY      814 ----- 813
Db      2340 KLTTTATTTSTGSTATPSSSQGPPAGTPAGPHVSTTATTPTVTSKATPSSPGTATAPAL 2399
      QY      814 -SEVSTPTTT-----KEPTTIHKS----- 831
Db      2400 ESTATPTATSTAIIPSSSLGTTWRLSQTTTPMATMSTATPSTPEIVHISTVLTTAT 2459
      QY      832 ----- 831
Db      2460 TTGATGSAVPSPSTPGTAHTTKVPTTTTGTFTVPSSSPGTA RTPVWISTTTTPTTSGS 2519
      QY      832 -----PDEST----- 836
Db      2520 TVTPSSIPGTHTPVLTTLTTTQVATGSMATPSSSTQTSCTGTPPSLIATTATTATATGTTN 2579
      QY      837 -----PELSAEPKALEN-----SPKEP- 855
Db      2580 PSSTPGTTIPPELITATTATPAATSVTPSSALGTHTPPVENVTTATHGRSLSPSSPH 2639
      QY      856 -----GVPTTKTPAATKPEMT----- 871
```

RESULT 6

T16251

hypothetical protein f35A5.1 - Caenorhabditis elegans

C.Species: Caenorhabditis elegans

C.Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C.Accession: T16251

| | | |
|------------|-----|---|
| D b | 641 | E PIS - KP KD TAPKAE NSPVVP - PTPVKNPVKKWKP PPWEDDDADAPAKPVSLEPEEK TP 697 |
| Q y | 661 | --KET APTTPKG -----T APTTLKEBP A--TT PKKZAPK-ELAP TT 696 |
| D b | 698 | V LAKA PT KPOSEAAAD PVG S SKDP KLAKA PVK RDESPMKAVIKDPAKTEVP AV 757 |
| Q y | 697 | TK- EPTSTTS D----K PAPTTPKGTA TP -KEP ----- 724 |
| D b | 758 | V KKPEPVAKS RD PSPKKA EP NSP-VVP PTPVKNPVKKWKP PWEDDDADAPAEVNV PEP 815 |
| Q y | 725 | -----A PTTPKEPA -----P TPPKGTA PT LKEB APT -PKKAPKE ----- 760 |
| D b | 816 | E BKT TV LAK TPV KPRDP SP KCAVAPAK PS TKTDAPPVSVKKP EV SKPEPSKKA EPSN 875 |

| | | | | |
|----|------|---|---|----------|
| QY | 761 | --LAPTTTGGT----- | -----STTSDDKAPRTTPKETAFT--TPKEP-----AP----- | TTTP 798 |
| DB | 876 | PVVPTPTVKVPPVKWKPPWEDDDDEPTBEVKKPSEPEKKTFLVAKGKPEKPKDAPKVAAPK 935 | | |
| QY | 799 | KKPAP--TTPTTP----- | -----PTTSEVSTPTTTTKGPTTIHKSP--DEST 836 | |
| DB | 936 | RDPSEKKAAPKEKPAKVAAPRDLSPKKAIPIPANTQEAAPTPTVKNPVKWKPPWEDDDE 995 | | |
| QY | 837 | PELSAEP-----TPKALENSPKPGVPTTK--TPAATKPEMTTTAKDKTTERDLRT 885 | | |
| DB | 996 | P---AEEVSAPEPEKKTPTVLAKKAPAKPRDPSPKAAPVAAPKPD-----1036 | | |
| QY | 886 | TPETTTAAPKMTKE----- | -----TATTTEKTTESKITATTTQVTSITTQD 926 | |
| DB | 1037 | -PKTIPEVPTPTVKNPVKWKPPWEDDDDESEPVSAPEPK-----1075 | | |
| QY | 927 | TTTPKITTTLTKTTLAPKVTTKTTITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 986 | | |
| DB | 1076 | -----KTPVLAKKAPTATKATKPDSEAAADPVSGPTSKDPLSKKAPVEKP-KPTTD 1125 | | |
| QY | 987 | PK---KPTSTKKPKTMVRKPK----- | -----TTTPPKMTSTM-----1018 | |
| DB | 1126 | PKDDKILKPSAKKPEKAPKPAKPKWKPPWDDDDDEPEADFTVPAPSKKPDTEFPADPLG 1185 | | |
| QY | 1019 | -----PELNPTSGRIAEAMLQTTTRNPQTPNSKLVENPKSEDAGGAEGETPHMLLRPHV 1072 | | |
| DB | 1186 | GPXTKDPKLNKKA-----PAEKTEKPKPKVSKPEPK-----1218 | | |
| QY | 1073 | FMEVPTPDMYDLPRVNPQIIINPMLSDETNICNGKPDVGLTTLRNGTLVAERGHYFWML 1132 | | |
| DB | 1219 | --PTEPPK---PAAPKK-----WKPP-----WED 1237 | | |
| QY | 1133 | SP-----FSPSPARRITEVWGISPIDTVTRCNCEGKTFFFKDSQVWRETFNDIKAG 1186 | | |
| DB | 1238 | DPDEPEADFTMAPKPDTE-----DPADSL-----G 1264 | | |
| QY | 1187 | YPKP 1190 | | |
| DB | 1265 | GPKP 1268 | | |

RESULT 7

T31108

cyst germination specific acidic repeat protein precursor - Phytophthora infestans

C/Species: Phytophthora infestans (potato late blight agent)

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004

C/Accession: T31108

R:Goernhardt, B.

submitted to the EMBL Data Library, April 1998

A/Reference number: Z20986

A/Accession: T31108

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1489 <GOE>

A/Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:G3851513; PID:G3851514; P

C/Genetics:

A/Gene: car90

Query Match 14.0%. Score 1020.2. DB 2. Length 1489.

Query Match 14.0%; Score 1020.2; DB 2: Length 1489:

Best Local Similarity 24.9%; Pred. No. 7e-14;
Matches 431; Conservative 69; Mismatches 501; Indels 728; Gaps 79;

QY 27 LSCGRCFBSFGRBC-----DCDAQCKYD--KCCPDYESFCAEVH-----68
Db 43 IPCSGVGAEPV--GTACPKAGDVATSDCQPYLLSYNGAVCVAPVDAECALIHDDMWGCEP 100
QY 69 -PTSPSSKKA-----78
Db 101 PKTYTSVAEATIAAYNGESSGWTGHDEVVQVGDEEBEIPARVNYDTTVDTPIGVNC 160
QY 79 -----PPPSG-----ASQIKSTTKRSPKPPNKKTKV 107
Db 161 VATETATQGHATGGKYDTPSTGTQGVNTIHYGSTTTEGVTGGYPTD-----AKV 216
QY 108 IESE-----111
Db 217 IDGETYLDYPTGITEIIEBGTTPGGYTTDGGTTGGYTTVDNTHETEGAGGYDAGR 276
QY 112 -----EITEHSVSNQSSSSSSSSST-----TWIKSSKNSANRE 152
Db 277 EYESTTPTVGYSTETEGQHVTCGYEFSDETEAFTEGTYVPRBEETAAPESEDITYAPRE 336
QY 153 LOKKLKVKONKKRKKKPTPKPPVVDACSGLDNG-----188
Db 337 V-----TPVAPTEKPPVDVEETTYTEETTYAPTKSETNAPTERMHYAHIEKP 383
QY 189 -DFKVT-----TPDTSTTHQNKVSTSP-KITAKPAINPRPSLPNSDTSKETSUTVNKETT 242
Db 384 CDTEVIMYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 438
QY 243 VETKETT-----TNKQSTDSKKTSAKETQSTKISAKOLAPTSKVLAKGTPKAETTK 299
Db 439 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 495
QY 300 GPALTTPKEPTTPPKPEASTPKPEPTTIKSAPTTPKEPAPT--TTKSAPTTPKEPAPT 358
Db 496 ASTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 554
QY 359 --TTKEPAPTTPKEPAPTTPKAPTTTKSAPT-----TPKEPAPTTPKAPAPT 406
Db 555 EEITYAPABETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABET 613
QY 407 PKPEAPTTPKEPAPTTPKEP-----AP---TTKEPAPTTPKEP-----AP 443
Db 614 PYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABETPYEPTTEETTYAP 673
QY 444 TAPKAPAPTTPKBPAPTTPKEPAPT--TTKEPAPTTPKEPAPT-----APT--TTKSAPT-- 491
Db 674 TEETTYAPTEETTYAPTEETTYAPTEETTYAPABETPYEPTTEETTYAPTEETTYAPTEET 733
QY 492 -----TTKEP-----APT--TTKSAPT--TPKEPSPTTTKP-----APTPKBPAP 529
Db 734 MYAPTEETTYGPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 793
QY 530 TTPKBPAPTTPKEPAPT--TPKEPAPTTPKBPAPTAPKEP-----APITPK 573
Db 794 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 853
QY 574 ETAPTTPKLTPTTPKEPAPT--TPKEP-----APTPEELAPT-----610
Db 854 TYAPTEKTYAPTEETTYAPTEETTYAPTEETTYAPABETPYEPTTEETTYAPTEETTYAPTEET 913
QY 611 -----TPPEPTPTTPEEP-----APPTPKAAAPNTPKEPAPT--TPKEPAPT 651
Db 914 TEETTYAPABETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABET 973
QY 652 PKEP-----APTPKAPTTPKGTAP-----TTLKEPAPTTPKPP-----AP 689
Db 974 PYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABETPYEPTTEETTYAP 1033
QY 690 KE---LAPT-----TTKEPTSTSDKPA-----PTTPKGTAPTTPKEPAPTTPK 730

Db 1034 TEETTYAPTEETTYASTEETTYAPTEETTYAPABETPYEPTTEETTYAPTEETTYAPTEET 1093
QY 731 EPAPPTPKGTAPT--TLKEPAPTTPKPP-----APKE---LAPT-----TTKGP 769
Db 1094 TYAPTEETTYAPTEETTYAPABETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1153
QY 770 TSTSDKBPAPTTPKAPT--TPKEPAPTTPKPP-----APTTPPTPTTSEVSTPT-- 820
Db 1154 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1211
QY 821 -TTKEPTTHKSPDESTBELSABPTKALENSKPEGVPPTTKTPAANKPEMTTAKDKTT 879
Db 1212 EETTYAPT--EETPYEPTTEETTYAPT-----EETPYEPTTEETTYAP-----T 1250
QY 880 ERDLRTPTTPTTAAPKMTKET-----ATT--TEKTESKIATTTQVTTST--TOD 926
Db 1251 EETTYAPTEETTYAP--TEETMYAPIDEETTYGTEETTYAPTEETTYAPTEETTYAPTEET 1308
QY 927 TT--PFKITT--LKTTLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTPKPQ 981
Db 1309 TTYEPTGETTYAPTEETTYAPTEETTYAPM-----EET-----PY 1343
QY 982 KPTKAPKPTSTKPP-KTWPVRVKKPTTPTRPKMTSTMPBLNPTSRIAEAMLQTTTPNQ 1040
Db 1344 EPAEESTSTVSTKPCNTBEFTDEPTDE-----PSDEPTD-----EPTDBPTD 1390
QY 1041 TPNSKLVEVNPXSDAGGAGETPHMLLRPHVFMPEVTPDMDYLPRVP--NOGIIINPML 1098
Db 1391 LPTE-----PSTPCDNQGI-----1405
QY 1099 SDETNICNGKPDGLATLRNGTLVAPRGHYFWMLSPPSPARRITEVWGI-----PSPI 1154
Db 1406 -----NLGIVENKYVYNA-----GLYNTTTPG--1428
QY 1155 DTVTRCNCBGKTFPFKDSQYWRFT-----NDIKAGYKPIPKGFG--CLTGOIYVLA 1207
Db 1429 -----RNSQSWHSCCRSCYND-----PICHAFSHQTSSDSVCLE 1464
QY 1208 TA-----KYKNWPESVYFFKRGSSIQQYIYKQEPVQKCPGRPALNYPVVGEMTVQVRR 1261
Db 1465 TSTSDREDOQNW-----LAGNMDR-----1484
QY 1262 RFERRAIGPSOHTIRIOYSPARLAYQKGLVHNEKVSILWRGLPNVVT 1310
Db 1485 -----NVVT 1488

RESULT 8

S48478

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)

N;Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019C

C;Species: Saccharomyces cerevisiae

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: S48478; A26877; B26877; S27281; JC6123

R;Rowley, K.

submitted to the EMBL Data Library, October 1994

A;Reference number: S48478

A;Molecule type: DNA

A;Residues: 1-1367 <ROW>

A;Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; G

J;Yamashita, I.; Nakamura, M.; Fukui, S.

J;Bacteriol. 169, 2142-2149, 1987

A;Title: Gene fusion is a possible mechanism underlying the evolution of STAL.

A;Reference number: A91831; MUID:87194600; PMID:3106330

A;Accession: A26877

A;Molecule type: DNA

A;Residues: 1-242 <YAM>

A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525

A;Accession: B26877

A;Molecule type: DNA

A;Residues: 762-1331 <YA2>

A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526

R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.

FEBS Lett. 239, 179-184, 1988

A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar

A;Reference number: S27281; MUID:89031230; PMID:3141213

A;Accession: S27281

A;Molecule type: DNA

A;Residues: 1-31 <PAR>

A;Cross-references: EMBL:X13857; NID:94551; PIDN:CAA32069.1; PID:94552

R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.

Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996

A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy

A;Reference number: JC6123; MUID:96323237; PMID:8710886

A;Accession: JC6123

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-1367 <LAM>

A;Cross-references: GB:U0626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387

C;Genetics:

A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458

A;Cross-references: MIPS:YIR019c; SGD:S0001458

A;Map position: 9R

C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein

F;5-21/Domain: transmembrane #status predicted <TM1>

F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 13.5%; Score 983; DB 1; Length 1367;

Best Local Similarity 24.0%; Pred. No. 3.6e-13;

Matches 377; Conservative 127; Mismatches 505; Indels 560; Gaps 71;

```

QY 9 YLL-----LLSVFVIQOVSSQELSCKGR-----FE----- 35
DB 118 YLIDNPTDFTAEVATQDVNS-----CQWMENFOIQEYLOGSAAQYASSNQWG 169
QY 36 --SFERGRCDDAQCKYKDKCPDYESFOAEVHNTPSPSSKAPPPGASQT----- 87
DB 170 TTSF-----DLSTGCNNYD-----QGHSTQDPFGFY 196
QY 88 -----IKSTTKRPPKPNKKTKKVISSBTEHVSSESSSSSSSSSSSSSSSSSS 136
DB 197 WNIDCNNCGYKSSITTSSTSSSTTSSTSSSTTSSTSSSTTSSTSSSTTSSTSS 256
QY 137 TIWKIKSSNKAANRELQKLVKDKNQRKTKKPTKPPVVDAGSLDNGDFKVTTPD 196
DB 257 TAPATPTTSC-----TKKPTP----- 274
QY 197 TSTQHNKYSTPKITAKDINPRSLPNSDTSKETSITVNKETTVEKETTNNKQTS 256
DB 275 -----PTTSCTKERP-----TPPHD-----TTPCTKKKTTTSK-TC 306
QY 257 TDGKEKITSAKETQSTKTSANDLAPTSKVLAKPTPKAETT--TKGPAITTPKE----- 308
DB 307 T---KKTTFVTPPS--SSTTESSAPV-----PTPSSSTTESSAPVTSSTTESSAPV 356
QY 309 PPTTPKEPASTTPKEPTPTTKSAP-----TTPKEPATTTKSAPTTTKEPATTTPKEA 364
DB 357 PTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESS 413
QY 365 PTPKEPATTTPKEPATTTKGAP-----TTPKEPATTTPKEPATTTPKEPATTTPKE 420
DB 414 -TESSAPVTS-----STTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPV 461
QY 421 ---TTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 477
DB 462 SSTTESSAPVTS-----SSTTESSAPV-----PTPS-----SSTTESSAPVTS 505
QY 478 PKEPATTTPKSAPTTTKEPATTTKSAPTTTKEPATTTPKEPATTTPKEPATTTPKE 537
DB 506 ESSAPVTPSSSTTESSAPATTSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV 565
QY 538 TTP-----KEPATTTPKEPATTTKK-----PAPT-----APKEPATTTPKE 578
DB 566 STPVTSTTESSAPVPT-----PSSSTTESSAPVPTPSSSTTESSAPVPTPSSST 621

```

```

QY 579 TPKKLTPTTPEKLAAPTTPKEPATTPEELAPTTPEETPTTPEEFAPTTPKAAAPNTPK 638
DB 622 SSAPVTSSTTESSAPVPTPSSSTTESSAPV-----PTPS-----SSTTESSAP-VTP 671
QY 639 PAPTTPKEPATTTPKEPATTTPKETAPTTPKGATPTTLKEPAP-----TTPKKAPKELAPT 695
DB 672 SSSTTESSAPVT-----SSTTESSAPVTSSTTESSAPVPTPSSSTTESSAP-----VPT 724
QY 696 TTKEPTSTTSDKPAAPTTPKGATPTTPKEPATTTPKEPATTTPKGATPTTLKEPAP-----TT 752
DB 725 ---PSSSTTESSAPVPTPSSSTTESSAPVT-----SSTTESSAPV-----PTPSSSTT 772
QY 753 PKKPAPKELAPTTTKGPTSTTSKPAAPTTPKEPATTTPKEPATTTPKEPATTTPKEPAPT 811
DB 773 ESSAP---VPT-----PSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTNITSSAPSS 825
QY 812 T-----TSEVSTPTTKEPTTIHKSP-DESTPELSAETPKALENSPKPEGVP----- 858
DB 826 TPFSSSTESSVPTPSSSTTESSAPVSSSTTESSAPVPTPSSSTNITSSAPSSIPF 885
QY 859 -----TTKTPAATK--PEMTTAKDKTTERDLRTTPTTAAPKMKETATTTEK 906
DB 886 SSTTESFSTGTTVTPSSSKYPGSGTETSVSSTT-----TTIVP-----TKTITSV 931
QY 907 TTESKITAATT-----QVTSSTTODTTPPKITTLKTTLAPKVTTT- 947
DB 932 TTPSTTTTITTVCGSTGNSAGETTSKPKVITVTPTTTTTSSTTITTTVCGSTG 991
QY 948 -----KKTITTT-EIMNKPEETAKPKDRATNSKATTPKQKP----- 983
DB 992 TNSAGETTSKPSKPTITTVPCSTSPSETA-----SESTTSTPTPTTVVSTTWT 1043
QY 984 -----TKAPKPTSTKPKTPMRVKPKTTPPKMTSTMPKMTSTMPKMTSTMPKMTST 1037
DB 1044 TEYSTSTKPGGEITTTFTVKNIPTTYLTIAPTP-----SVTTV 1082
QY 1038 PNOTNSKLVEV-----NPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLRVPNQGI 1092
DB 1083 TNFTPTTITTTVCSTGNSAGETTSKPSKPTVTTTVPSTGTGEGYTEATTLVTTAVTT 1142
QY 1093 IINPMLSDETNICNGKPDGLTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPS 1152
DB 1143 VTTESSTGNSA-GATTTGYTKSVPTT-----YVTLAPSAVPTPATN-----AVPT 1190
QY 1153 PIDTVFTRCN-----CEGKTFEFDKQSYWRFNDIKDAGYKPKPIKGFGLTG 1200
DB 1191 TITT--TECSAATNAAGETTSVCSAKT----- 1215
QY 1201 QIVAAIS-----TAKYKNWP-----ESVYFFKRGSGIOQYIYKQEP-----VOK 1239
DB 1216 -IVSSASAGENTAPSATPTPTTAIPTTITTESSVGTNSAGETTTGYTKSIPTTITTL 1274
QY 1240 CPGRRPALNYPVYGWT-----QVRRRFEAIGPSQTHTIRIQYSPARLAYQDKG 1290
DB 1275 IFCSGAKNYETVATNPDISIKTSQATTAASASVAP----- 1313
QY 1291 VLHNEVKVSIILWRGLPNVVTSAISLPNKRKPDGYDYAFSKQYXNIDVPSRTARAITTR 1350
DB 1314 -----VWTS-----PSLTGP-----LOSASGSAVATY 1335
QY 1351 SQQTLSKWW 1359
DB 1336 SVPSISSTY 1344

```

RESULT 9

T29757

protein UNC-89 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C;Accession: T29757

R;Du, Z.; Le, T.T.; Wilson, R.


```

QY 786 -----APTPK----- 791
Db 2443 NPVSKKRDVQLAVKKVGDAFTFAKNLEDRLITEGELTMDAKNIWKPKITWLKXDV 2502
QY 792 ----- 791
Db 2503 EITSDGHYKIVEEDGSLKLSILQTKLEDKGRITIKAESEBFGVAECASLGVVGRPKMAK 2562
QY 792 -----EPAPTT----- 797
Db 2563 PAFQSDIAPINLTGDTLECKLLITGDTFPFKVYIGTQLVCATEDTEISNANGVYTMKI 2622
QY 798 -----PKK-----PAPTPPET----- 808
Db 2623 HGVYADMTGKIKVAYNKAGEVSTEGPLKVAPIPEVFETSLCDATCREGDTLKLRAVLL 2682
QY 809 --PPPTTS-----EVSTPTTTKEPT----- 827
Db 2683 GEPEPVSVYVNGKLEESQNIKHSEKGTVTVIKDICTDYSQVVCEALNEVYKATSE 2742
QY 828 -----IHK----- 830
Db 2743 ATLLVLPGRPPDFLEWLSNVRARTGTVVHKVVFVGTGDPKPSLTWYINNKEILNSDLYTI 2802
QY 831 -----SPD-----ESTPELSAE-- 842
Db 2803 VTDDKTSTLTINSFNPDPVHVEIICKAENDAGEVSCVANMITYTSDMFSESESAQAEEF 2862
QY 843 -----PTPKALENSPK-----EPGVPTTK----- 861
Db 2863 VGDLDTEDESLREMHRTPTPVM--APKFIITKIDTKAKKGHSFAVFCVVPDTKGVCCK 2919
QY 862 -----TPAAT-- 866
Db 2920 WLKDGKEIELIARIVQTRTGPPEGHITQELVLNDVTPEDAGKYTCIVENTAGKDTCEATL 2979
QY 867 -----KPMETTTAKDKTTERDLRTT-----PETTTAAPKMTKETATT 903
Db 2980 TWIESLEKKSEKAPFEIVALQDKTKTSEKVLVECKVGEPPKPKVSWLHDNVSRKNPN 3039
QY 904 TTKT-TESKITANTTQ--VTSNTTQDTPPKITTL-----KTTTLAPKVTITTKT----- 950
Db 3040 SEKITQESITVBSVEGVERVITTSSE-----LSHGQKYCIAENTEGTSKTEAPL 3090
QY 951 -----ITTEIMNKP-----BETAKPKDR----- 969
Db 3091 TVQGEAPVFTKELQNKELSIGELVLSCSVKGSQPQPHVDYFSFSETTKVETKITSSRIA 3150
QY 970 -----ATNSKATTPKPKPT--KAP-----KKPTST 993
Db 3151 IEHDQNTHWRWISQITKEDIVSYKAIATNSIGTATSISKITTKVEAPVFEQGLKKTIV 3210
QY 994 K-----KP-----KTMPR-----VRKPKTT----- 1008
Db 3211 KEKEIKMEVKVGGAPDVEWFKDDPVEDSGDNHEMKKNPETGCVFTLVVKQAATTIDAGKY 3270
QY 1009 -----PT-PRKWTISTWPELNPITSRIAEAMLQTTTR----- 1037
Db 3271 TAKASNPACTAESSAAEAVTQSLEKPTFVRELVTTEVKINET-----ATLSVTVKGVDPD 3325
QY 1038 -----PNQTPNSKLV-----EVNPKSEDAGAGS 1061
Db 3326 SVEWLKDGQPVQTDSSHVITAKVEGSGSYITIKDARLEDGKYACRATNP-----AGEAKT 3381
QY 1062 ETPHMLLR---PHVFMPEVTP-----DMD----- 1082
Db 3382 EANFAVKNLVPPEFVEKLSPLEVKEKSTTILSVKVGVPFSPVFWFKDDTTPISIDNVHV 3441
QY 1083 -----YLPVPVNO-----GII-----INPM 1097
Db 3442 IOKQTAGFSLTINDAQGDVGVISCRARNEAGEALTATNFGIIRDSIPPEFTQKLRLP 3501

```

RESULT 10

T34513

hypothetical protein ZK783.1 - Caenorhabditis elegans
C.Species: Caenorhabditis elegans

```

QY 1098 -----LSDETNI CNKPKVD-----GLTTLRNGT 1120
Db 3502 EVREQETLDLKVTVIGTPVNPVWFKDD-----KPINIDNSHIFAKDEGSGHHTL--T 3552
QY 1121 LVAFRGH-----YFWMLSPF--SPSPARRITE 1146
Db 3553 IKQARGEDGVYTCATNEAGEAKTTANMAVQEBIEAPLFVQGLKPKYEVBOGKPAELVVR 3612
QY 1147 VMGIPSP----- 1154
Db 3613 VEGKPEPEVKWFKDGVPIAIDNQHVIEKKGENSHTLVIKDTNNADPKGYTCQATNKAGK 3672
QY 1155 -----DVFTRCNCCEGKTF-----PFKD 1172
Db 3673 DETVGLKIPKYSFEKQTAEEVKPLFTEPLKETFAVEGDTVWLECKVNKESHPOIKPFKN 3732
QY 1173 SQYWR-----TN-DIK----- 1183
Db 3733 DQPVETGOHMQLEVLEDGNIKLTIQNAKEDVGAYRCEAVNVAGKANTNADLKIQFAAKV 3792
QY 1184 -----DAGYKPKIF----- 1192
Db 3793 EEHVTDESGQLEBIGQFETVGDTPASSKTDTGRGAPFVELLRSCVTVTEKQQAIIKCKVK 3852
QY 1193 -----KGFGLT----- 1199
Db 3853 EPRPKIKWTKEGKEVMSARVRAEHKDDGTLTLTFDNVTQADAGEYRCEAENEYGSATWE 3912
QY 1200 GQIVAAALSTA-----KYKWPE-SVYFFKRGSGIQ- 1228
Db 3913 GPIIVTLEGAPKIDGAPDFLOPVKPAVVTVGETAVLEGKISGKPKSPKWKYKNGBELPK 3972
QY 1229 -----QY-----YKOBPVQKCPGRRP 1245
Db 3973 SDRVKIENLDDGTQRLTVTNAKLDDMDREYRCEASNEFGDVMSDVTLITKPEPAVFG-- 4029
QY 1246 ALNYPVYGEWTVRRRRPFA----- 1266
Db 4030 -----PFKLSAIQVKTETETAKFECKVSGTKPDVKWFKDGTPLKEDKRVHFESTDDGTQR 4084
QY 1267 --IGPSQTH--TIRIQSPARLAYODKGVLLHNEVKVSI-----LWRGLPNV----- 1308
Db 4085 LVIEDSKTDDQGNRYLEVS-----NDAGVANSKVPFLTVVPSETLKKIKGLTDVNVVTQGT 4138
QY 1309 -----VTSA----- 1312
Db 4139 KILLSVEVEGKPKTVKWKGTETVITSSQTTKIVQVTESEYKLEIESAEMSDDTGAYRVVLS 4198
QY 1313 -----ISLPNIRK----- 1320
Db 4199 TDSFSVSSATVTVTKAAEKISLPSFKGLADQSVFKGTPLVLEVEIEGKPKDKVKWKNG 4258
QY 1321 -----PD----- 1322
Db 4259 DEIKQKVEDLNGKXRLTIPFOEKDQVGEYSVTAANEAGEIESKAKVNVSAKPEIVSGL 4318
QY 1323 -----GYDYAFSK-----DQYINIDVP----- 1340
Db 4319 VPTTVKQGETATFNKVKGVKVKYKNGKEIPDAKTKDNGDGGSYSLIPEIPNAQVEDAAD 4378
QY 1341 -----SRTARAI-----TTRSGQTLIS--KV----- 1358
Db 4379 YKVVVNDAGDADSAALTVKLADGDKVKVPEIVSGLIPTTVKQGETATFNKVKGPVK 4438
QY 1359 ---WY 1360
Db 4439 QVKWY 4443

```

| | | | | | | | | | |
|----|------|---|-----|-----|-----|-----|-----|-----|-----|
| QY | 19 | IQQVSSQBELSKCKRCFSFE-RGRECDCA-----OQKXDKC----- | : : | : : | : : | : : | : : | : : | : : |
| Db | 506 | VELTTSGLACTSCPPNSECVGVCVSGYNALVGCEDIDECITEICNIEANWCW | : | : | : | : | : | : | : |
| QY | 56 | -----CPDY- | : | : | : | : | : | : | : |
| Db | 566 | NLIGFVCCNPTNATHDCDIFLKVKIYAWIIIFLLKGLEITKEXLHVIGNEBDTV | : | : | : | : | : | : | : |
| QY | 60 | -----ESFCAEVH---- | : | : | : | : | : | : | : |
| Db | 626 | VATRSNHSTSDQLITVVQQRNFSTGQIILTRGVSSGEAVTOTTDADFGLEISAADL | : | : | : | : | : | : | : |
| QY | 68 | -----NPTSPSSKKAPPPGSA-----SQTIKST-- | : | : | : | : | : | : | : |
| Db | 686 | AGSGSIGTLPTTLEPKIEGGKKA---SGGYMTBEDEGEDLMEEGSGSWTTINGFI | : | : | : | : | : | : | : |
| QY | 92 | -----TKRSPKNKKTKKVIES---EELTEHSVSFN | : | : | : | : | : | : | : |
| Db | 743 | TGSPRSEGTIRVRITTLGEDGETATRKGISADPKTGEGSKTESDGEEKLTVDKGKEA | : | : | : | : | : | : | : |
| QY | 123 | QESSSSSSSSS | : | : | : | : | : | : | : |
| Db | 803 | QSSGSATSCKKSEATSGSSSSSAKSCTGSEAGSSGASSSSSGSVSGSGSVSTBSG | : | : | : | : | : | : | : |
| QY | 133 | -----SSSSTTWIKISSKN----- | : | : | : | : | : | : | : |
| Db | 863 | SFGTSSSGSVGSEATGSTGDGSESGKPSSKEKLPTKNGEKSPISGSDTTTKESSE | : | : | : | : | : | : | : |
| QY | 147 | -----GAANRELQKKLV----- | : | : | : | : | : | : | : |
| Db | 923 | ETTSRKPIEGDSLTEGSGSGHWFTGSKGHFSGSKSVSVTGKPTOSGAEAGSGSGPKV | : | : | : | : | : | : | : |
| QY | 160 | -----KONK----- | : | : | : | : | : | : | : |
| Db | 983 | PXGPAGEPITTDGBESSSTSTGDKSGKPADKSDNKNPVKTDKNPDITTDGEDSTSETS | : | : | : | : | : | : | : |
| QY | 164 | -----KNRTK-----KKTP-----KPVVDEAGSGLDN | : | : | : | : | : | : | : |
| Db | 1043 | GGEQGPKGSKQPQDPDKGSEVKKPTSEVDGFGNLSGTGKSNVFLKPTDLPEEGSGI- | : | : | : | : | : | : | : |
| QY | 188 | GDFKVTPD---TSTTQH-----NKVSTSPKITTAKPINPRPSLPNPNDSKE | : | : | : | : | : | : | : |
| Db | 1101 | ----LTTSSGGKNSTFEHTKLERLPKPTEDKSSFTPOLGLEISAKKPE--PEDGTSKE | : | : | : | : | : | : | : |
| QY | 233 | TSLTVNKETTVETKETTTTTNKOT-----STDGKEKTTTSCKET | : | : | : | : | : | : | : |
| Db | 1155 | VGLEILLWEST--TPGSTTILDSVGLIEISGSDLTKATKKPHVEIBSGTGDBEITAT-T | : | : | : | : | : | : | : |
| QY | 270 | QSIEKTSAK-----DIAPTISKVLAKTTPKPAETTITKGPALTITKEPT- | : | : | : | : | : | : | : |
| Db | 1211 | RDVSKSTKPRVEVDGDNGETSGVDGKPTTTPATFPSSAESSTGRIPITTSASPEGSGG | : | : | : | : | : | : | : |

```

QY 757 APKEL---APTITKGPST---STT-----SDKP----- 777
Db 2321 SEVLSNBPSTTEAPTTLSPDILSTTTNNLSQSSTVSTEDSEISENSEKPTSPAPLV 2380
QY 778 -----APTTP-----KETAPTP-----KEPAPTT 797
Db 2381 TSSVTHVASSGPDVPTESSEPDLTGSSTENIPEASSKQTISSTPTDPTTASSEPTKST 2440
QY 798 PKXP-----APTTP----- 807
Db 2441 SMSFDLSTSNVLSSESTTPSSSKSPVSSSTEGISVVSTFEFKVPSTISSVLEEDLT 2500
QY 808 --TPPTTSVSTPTTKEPTT-----IHK--SPDESTPELSAPRTPKALENSP-K 853
Db 2501 KTTSPILEETTTASSETSEPLETDSLTVSVRIHELTTSSENVPKESBTTTSSSSKPSQ 2560
QY 854 EPG-----VPTT-----KTPAATKPEM-----TTTAK 875
Db 2561 EPAGILSTVVVPTSSVSLITASEIAITSNTPFKQGRTPITTPSKLVKSTTSPSTVTS 2620
QY 876 DKITERDLRTPTTAAPKMTKETATT-----TEKTESKITATTTQVTS-- 921
Db 2621 SEPESEIKRTVSTVSTTPTTEFTTSESILITAAPSKPTESITESBAPTTPAKTSET 2680
QY 922 -----TTTQDTTPFKITTLKTT-----TLAPKVTTTTKTTITTT----- 954
Db 2681 KPSNVSTSRKSTENVETSTSQSGLESSTMSSTSSPEINAPATVVSSEASSTLEENS 2740
QY 955 -----EIMNKPEE----- 962
Db 2741 STSPTSSEASVKLSLFPESITSEAVTVSSRAPEITMSSESHREISTVSSSEPEIP 2800
QY 963 -----TAKPKD-----RATNSK 974
Db 2801 LSTTVSENVVTASSIPSEEPILSVSTSSSTPRVRLITGTDDLLVSVTVFSGHNRNRNIT 2860
QY 975 A-----TTPKQKPTKAPKPTSTPKPKTWPVRVKPK-----TTTPP-- 1011
Db 2861 ASSVPSNSTSPIILPSESITTPQPPPTTTTAKPATTSGRGPSPSIQPPAEMTTTAPPP 2920
QY 1012 -----RKWTSMPELNPFSRIAEMLQTTT----- 1036
Db 2921 PSNGGYEETNOBEEQVST-----TTTEAPSLCSTVTCHSLATCEQSTGVCICRDGFI 2974
QY 1037 -----RPNQTESKLIVEVNPXSEADAGAGEHTPHMLR 1069
Db 2975 GDGTTACKSKSTADCLSLPSLCAKADKCNSTRSCEC-----DAG----- 3014
QY 1070 PHVFMPEVTPDMYLRVNPQGIINPMLSD-----ETNINCKREV-----DG 1112
Db 3015 -----YI-----GDGYVCSFHPQDCVLRDNLCSPEAVCQNRRCQCLPGFTGDG 3057
QY 1113 LTTLR-----NGTLVAFRGHYFWMLSPLFPSPFPSPARRITEVWG----- 1149
Db 3058 VKCVSITHERASNCSCDANNAHCVGGTCKCNPGYF-----GNGLCC 3098
QY 1150 IPSPIDTV-FT-----RCNCEGKGTFFFKDSQYWFRTNDIKDAGYKPIFG 1194
Db 3099 VFDELDCVHTGTICHPNVNCNSESRCQCS-----GSCICRHGFKGDPFY 3175
QY 1195 FGG-----LTGQIVAA-----LSTAKYKNWPESVYFFFKGSGSIQOYIKQEPVQ 1238
Db 3131 FSGNGVSCFPKSCRTCDKSVCAKNAICLPT-----GSCICRHGFKGDPFY 3175
QY 1239 KCP---GRBPA-----LNPVYVGEWTVQVRRRRRPER--ATGPSQT--- 1272
Db 3176 KCSSLVSKFEPANQODLSVDSSVCPDCSSTQLCISGEC--ICKSGFRNRNSTLSGSETCAD 3233
QY 1273 -----HTIRQYSPARLAVQDKGV 1291
Db 3234 IDECAEKSHKDRVATCRNTFGSHVCTCPDGHVGDIITCVPH----- 3276
QY 1292 LHNEVKVSLWR--GL-----PNVVTSAISLPIRKP----- 1321

```

```

Db 3277 --NOKUSVYCEADGTMVLVGNETADFGKIFVKGQAEINPYCSKSFSSLLNSRKPYPVKV 3334
QY 1322 -----DGVDY----- 1326
Db 3335 VFQCHDVQLLDNHTWASTVVVVKHAMFLTNKADSYDLRCQPIGSRVESHVNVSELAATT 3394
QY 1327 -----YAFSKDOYNNID-----VPS-----RTARAIATR 1350
Db 3395 STLTKNSTLAPICRLSVSNDOHSSISSAMVGVDTLKLALEVTPTSENFOLLPRNCFAINIE 3454
QY 1351 SGQ 1353
Db 3455 SGE 3457

RESULT 11
A41819
proline-rich peptides 637K precursor, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 09-Jul-2004
C:Accession: A42663; A41819; A31966; B20593; A20593
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the
le exon.
A:Reference number: A42663; MUID:92250652; PMID:1577819
A:Accession: A42663
A:Molecule type: DNA; mRNA
A:Residues: 1-3706, 'I', 3708-4077, 'F', 4079-4155, 'S', 4157-5761 <DE2>
A:Cross-references: UNIPROT:Q63455; GB:M86514
A:Experimental source: ventral prostate
A:Note: sequence inconsistent with the nucleotide translation
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
Submitted to the Protein Sequence Database, April 1992
A:Reference number: A41819
A:Accession: A41819
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-5762 <DE1>
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Wind
J. Biol. Chem. 263, 19159-19165, 1988
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich p
A:Reference number: A31966; MUID:89066721; PMID:3198617
A:Accession: A31966
A:Molecule type: mRNA
A:Residues: 3372-3540 <HEM>
A:Cross-references: GB:M20721; GB:J04188; NID:g206397; PIDN:AAA41950.1; PID:g554494
R:Peeters, B.; Heyns, W.; Bosseyns, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary s
A:Reference number: A94675; MUID:84061859; PMID:6685733
A:Accession: B20593
A:Molecule type: protein
A:Residues: 2020-2057 <PEB>
A:Note: this peptide, designated proline-rich polypeptide V, can be found at several loc
A:Accession: A20593
A:Molecule type: protein
A:Residues: 2822-2859 <PE2>
A:Note: this peptide, designated proline-rich polypeptide IV, can be found at several lo
C:Genetics:
A:Introns: #status absent
A:Note: single copy gene with no introns
C:Superfamily: rat prostatic proline-rich peptides 637K precursor
C:Keywords: prostate; tandem repeat
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <MA

Query Match 12.6%; Score 919.5; DB 2; Length 5762;
Best Local Similarity 11.4%; Pred No. 1.8e-10;
Matches 526; Conservative 196; Mismatches 474; Indels 3404; Gaps 150;

```

```
QY 35 ESFERGECDDAQCKKYDKCDDY-----ESFCAEVH-NFTSPSSKKAPP 80
Db 850 ETTYSKQVDSLSTHLKTSLELPPDYTWGLEPSLYQQITQSPPKSLHENPKSF-----902
QY 81 PSGASOTIATTKSPKPNKKTKKV-----ISEBEITEHSVSEN 122
Db 903 -----VLKPSAAQALKPRKEKLSPTNNMI PHHPKPKLKNMWHI PAHKMTVPRQIQED 956
QY 123 QESSSSSSSSSS-----SSSTIWKIKS-SKNSAANRELQKLVKDNKNKRTKK-- 169
Db 957 QGYTIISSNGSFQPLDLEVLITSGIIPVVKHILPKRTVNPQTVSYQVKISHSQHVETQHPN 1016
QY 170 -----KPTPKPPVDEAGSLDNGDFKVTTPDTST-----199
Db 1017 SETTVQPLDLEFAINLQPTPKE-----NFAQTFQDTTQMIGPCKEVIAQA 1063
QY 200 -----TOH 202
Db 1064 PEHHEGTIPIIQOABYSTLPVTSFQPLDQELITSEAIREFPHPTVPOQTIIVHPPH 1123
QY 203 NKVSTGPKITAKP-----INRPS-----222
Db 1124 PLVHSEQVYQHPNTEFAIIQPLDLELITPQPTAEGELPQTLQDSTTQIIEBPTVVVG 1183
QY 223 -----LPPN--SDTSKETSIVNKETTVET-----KETTTN-- 252
Db 1184 PVPIYEBVTQTTSDQOAEYPPSPVTSFQSLDLELITSPATRESYHPSLLQQTIVNPP 1243
QY 253 -----KOTSTD-----GREKTTSAKETQIEKTSIAKDLAPTSKV 286
Db 1244 EHPVLVHSEQVHTQHPNLTEATVQPLDPLDITPQPTTEGELPQTLQDS-----TSQI 1296
QY 287 LAKPTPK-----AETTTKGPAITTP-----KEP-TPTTPKEPA 318
Db 1297 IIEPPIKVALPVYQVQSQAQAEVTTSTVSFQPLDQELITSEAIREFPHPTVPOQTI 1356
QY 319 STTP-----KEPTPTTKSAP-----TTPKEPA---PTTKSAPITTPKEPA 356
Db 1357 TVHPTKPLVTHSRQTHPNTTEVTVQPLDLELMTPTPTAAGELPQTLQDSTTQIIEP- 1415
QY 357 PTITKEPAP-----TTPKE-----PAPTT-----375
Db 1416 PTVVVGVPYIEEVTVQTTSDQAEYPPSPVTSFQSLDGLITPTPEHEITQKTIVFP 1475
QY 376 -----TKBPAPT-----TTPKAPTTKEP-----394
Db 1476 PMYTDVTLPOQVSVQHLKPTBEGIVQPLDLELITPQPTPEGELSQTQVESTQNKEPHE 1535
QY 395 --APT-----TPKK-----401
Db 1536 VVAPVPVYQAVTVPTFSQYQAEYQKSLQPLDLELTVTSEPTKEAVHSTISKNLAINQY 1595
QY 402 -----PAPTT-----PKE-----409
Db 1596 VHIQHNPDAEATVQPLDLELITSSSQPTAEGELLYSMQETVQIASEPPKQVWTVVPEYQ 1655
QY 410 -----PAP-----TTPKEPTPTTPKE-----425
Db 1656 EVAVFAPYQDAQKYPVLSIVSLNSLDQELITLSSELLGEAHLITPDETWLWLPKDRQGIY 1715
QY 426 -----425
Db 1716 PDHDKHLNLTETVNPQPHLEHTVQHOFTIEBERSQSIQKTTQITPTEPGKVVVLAQES 1775
QY 426 -----PAPTTKEPAPTTP-----438
Db 1776 BEVTIPMLKETAPPTPHSMALQSLDEKLITIHSHSPGWTQOAHANLKSOGHTGKILLD 1835
QY 439 -----438
Db 1836 YAEPNMEIELKHGLFLKTTBATESNTNQMTKSLKQVTLFTQNKKSMLPALVESQD 1895
QY 439 -KEPAPTAPKP-----APTTPK-----EPAPT-----460
Db 1895
QY 1896 ESQPPNMNSLQPLDQELITSSQPHGWPVHPINTEPKIYLHYAEP-PTGPEVEPPELFFLK 1954
QY 461 -----TPKE-----PAP-----TT 469
Db 1955 TTKSRPQVQTATQMAASPKEMVSRAPENKEAVLSGPGEDQDESPSPNMNSLQSDOELTL 2014
QY 470 TKEPSPTTTPKEP-----APT-----TTKSAP-----490
Db 2015 SSOPHGWIPHPNTHGKIYLHYAEPPTGPFVEPPDLFFLKTYSKPVWVWTLTRTKSRKE 2074
QY 491 -----TTTKEP-----496
Db 2075 MVSQSPKYEEAVLVHGEQESRSPNMNSLQSLQELITSSQPHGWPVHPNTHGKIYL 2134
QY 497 ---APT-----TTKSAP-----505
Db 2135 HYAEPPTGPFVEPPDLFFLRTTKSKPVQGTATRMVKSPEBMVSLDENKEAVTFAQGEK 2194
QY 506 -----505
Db 2195 GESPSPPNMNSLQSLDHELFMSQPHGWI PHHPKTPDKIYLHYAEPPTGPFVEPPDLFFLR 2254
QY 506 TTPKEPSPTTTPKEPAPTTP-----KKPAP-----537
Db 2255 TTKSKPVQGTTEMA-KSPKEMVSTQPEYKEAVLSGPGEDQDESPSPNTSLKSLDOEVA 2313
QY 538 --TTPKEPAPTTPK-----EPAP-----TTPK-KPAPTAP-----564
Db 2314 MSSQPHSGVPHHPKTPGKIYLHSEPPPGPFVKPTDILVKTITTKSKPAEWTPRIDKLL 2373
QY 565 KEPAPTTPK-ETA-----TTPKKL 583
Db 2374 KEMVHSGPEYEEAVFPFAHGQDESGSPNMPLQPLDQELITSSQPHGWPVHPNTHGTEGKI 2433
QY 584 -----TPTTPEKLIAP-----TTPKEPAPTTPBELAPTTPBEPTTPP-----620
Db 2434 YLHYAEPPTGPFVEPPDLFFLRTTKSKPVQSPKXIA-KSPKEMVSTQPEYKEAVLSQGP 2492
QY 621 ---EBAP-----TTPKAAAPNTPK-----EPAP-----641
Db 2493 EDQDESPSPNMNSLKLSDQELITSSQPHSGVPHHPKTPGKIYLHSEPPPGPFVKPTDLI 2552
QY 642 -----TTPKEPAPTTP-----KEPAPTTPK-ETA-----664
Db 2553 LVKTTTKSKPAEWTTPRIDKLLKEMVHSPSEYEEAVFPFAHGQDESGSPNMPLQPLDQ 2612
QY 665 -----PTTP-----KGTAPTTL- 676
Db 2613 ELTSSQPHGWPVHPNTPGKIYLHYAEPPTGPFVEPPDLFFLRTPKSKPVQGT-PTQMA 2671
QY 677 -----KEPAP-----681
Db 2672 KSPBEMVSLSPKNKETVFPFAQKGQDESPSPNISLQSDQELITMSSQPHGWI PHPPNTH 2731
QY 682 -----TTPPKP-----APKEL-----692
Db 2732 GKILHYAEPPTGPFVEPPDLFFLKTYSKPVQSPROI DKSPEMFTQSPYEESLLPA 2791
QY 693 -----AP-----694
Db 2792 HAEGQESRAPPHFSLQPLDQELISLSHPHGWIPHPNTPDKIYLHYAEPPTGPFVEPPD 2851
QY 695 -----TTTKEP---TSTTSDK-----PAP-----710
Db 2852 LFFLKTYSKPVQGTATKTDKSPEDRVSTQPEYKEAVLSGPGEDQDESPSPNMNSLQSLD 2911
QY 711 -----TTPKGTAPTTPKEP-----APT---727
Db 2912 QELATSSQPHGWI PHSPNAPDKIYLHYAEPPTGPFVEPPDLFFLKTYSKPVQGTPTQMA 2971
QY 728 -TTPKEPAPTTPK-----738
```

Db 2972 KSPKEMVQTPYKEYKADLSAPGENRDESPSPNMSLHPDQELSLSSQPHGWIHPHPPNTH 3031
Qy 739 -----GTAP-----TTLKEP----- 748
Db 3032 GKILYHYAEPPTGPFVEPPDLFFLTKTKSPVQSGPSQIDKSPKEVFSQSESESVLPA 3091
Qy 749 -----AP----- 750
Db 3092 OAEGQESRAPPHMSLQDQLTLSSHHPHGWIPHHNTPGKIYLYHYAEPPTGPFVEPPD 3151
Qy 751 -----TTPKXP-----APKEL-----APTTKG----- 768
Db 3152 LFFLTKTKSPVQSGPRQVDSKPKEMFTQSPYEESVILPAQAEQOESRAPPHMSLQPLD 3211
Qy 769 -----LFFLTKTKSPVQSGPRQVDSKPKEMFTQSPYEESVILPAQAEQOESRAPPHMSLQPLD 3211
Db 3212 QDLTLSSHHPHGWIPHHNTPDKIYLYHYAEPPTGPFVEPPDLFFLTKTKSPAQWTPQTOMA 3271
Qy 783 -----KETA----- 786
Db 3272 KSPPEMVSLSPENKETVPPAQCGQOESISPPHMSLQDQLTLSSHHPHGWIPHHNTPH 3331
Qy 787 -----PT-----TPKE-----PA 794
Db 3332 GKILYHYAEPPTGPFVEPPDLFFLTKTKSPVQSPRQIDKSPKEVFTQSPYEESVILPA 3391
Qy 795 ----- 794
Db 3392 OAEGQESRAPPHMSLQDQLTLSSHHPHGWIPHHNTHGKIYLYHYAEPPTGPFVEPPD 3451
Qy 795 -----PTTPKXPA----- 802
Db 3452 LFFLTKTKSPVHSGPRQIDKSHKEMFTQSPYEESVILPAQAEQOESRAPPHMSLQPLD 3511
Qy 803 -----PTTPTLP-----PPTSEVSTP-----TTTK-----EPTTIH 829
Db 3512 QDLTLSSHHPHGWIPHHNTPDKIYLYHYAEPPTGPFVEPPDLFFLTKTKSPVQSGPRQID 3571
Qy 830 KSPDE-----STPELSAEP----- 843
Db 3572 KSPKEMFTQSPYEESVILPAQAEQOESRAPPHMSLQDQLTLSSHHPHGWIPHHNTP 3631
Qy 844 -----TPKALENSPKX----- 854
Db 3632 DKILYHYAEPPTGPFVEPPDLFFLTKTKSPVHSGPRQIDKSPKEVFTQSPYEESVILPA 3691
Qy 855 -----PGVP----- 858
Db 3692 QAEDQESRAPPHMSLQPLDQDLTLSSHHPHGWIPHHNTPDKIYLYHYAEPPTGPFVEPPD 3751
Qy 859 -----TTKT-----PA----- 864
Db 3752 LFFLTKTKSPVHSGPRQIDKSPKEVFTQSPYEESVILPAQAEQOESRAPPHMSLQPLD 3811
Qy 865 -----ATKP-BMTTAKD 876
Db 3812 QDLTLSSHHPHGWIPHHNTHGKIYLYHYAEPPTGPFVEPPDLFFLTKTKSPVQWTPQTID 3871
Qy 877 KTERDLRTTPE----- 888
Db 3872 KSREEMVQSPENEEADIPRHGQDQELRNPIISLQPLDQELTLSSQPHGWIPHPSTP 3931
Qy 889 -----TTTAAKMTKETATTE----- 905
Db 3932 DKILYHYAEPPTGPFVEPPDLFFLRTTKSPVQGTPTQIAKSPKEMVFTQPEYKEAVLSA 3991
Qy 906 ----- 905
Db 3992 PGEDQDETSPSPNMSLQDQELTMSSQPHGWLPHPNTHGKIYLYHYADPPTGPFVEPPD 4051
Qy 906 -----KTTESK-LFATTTQVTSITTO----- 925
Db 4052 LFFLTKTKSPVQGTVTQMDKSPKQGLLPAHGEAQDESPPNMSLQPDQELSLSSQPH 4111

Qy 926 -----DTTPPKITT-----LXTT----- 938
Db 4112 GMITHPNTPGKIYLYHYAEPPTGPFVEPPDLFFLTKTKSPKLAGTPTQMAKSPKEMFSQT 4171
Qy 939 -----TLAP----- 942
Db 4172 LEHKEAVLSAPGEQODESPSSNMSLQSDQDITLSPOQYGRIPHVPVTPSGKIYLYHYVEH 4231
Qy 943 -----KVTITTK-TITTTIMNKPEE----- 962
Db 4232 PTVPLFQHPNMFSTKTTTSKPLQLTQTOMAKSPBEIVSLRPEYKEAILPAQVESOESEV 4291
Qy 963 -----TAKPK 967
Db 4292 LPNMSLQSDQELTLSSQPHGWIHPHNPDPDKIYLYHYAEPPTGPFVEPPDLFFLTKTKSK 4351
Qy 968 D----- 968
Db 4352 PVQWPTQIDKSLKQMAQSPYEBAVPPAHGEGODESRYAPNISLQSLDHDITLSSQPH 4411
Qy 969 -----RATNSKAT--TPKP----- 980
Db 4412 GMITHPNTPDKIYLYHYAEPPTGPFVEPPDLFFLTKTKSPAQWTPPTMAKSPPEMVSL 4471
Qy 981 ----- 982
Db 4472 LKNKEAVLPAQCGDEEESISPPNMSLQDQELTLSSQPHGWHHPNTHGKIYLYHYAEP 4531
Qy 983 PT-----KAPKK----- 989
Db 4532 PTGPFVEPPDLFFLTKTKSPVQGTITQMVKSPKQGLLPAHGEAQDESPPNMSLOPVD 4591
Qy 990 -----PTST-----KKPKTPRVRKP-----KT-----TPT----- 1010
Db 4592 QELSLSSQPHGWIPHHNTPDKIYLYHYAEPPTGPFVEPPDLFFLTKTKSKPLLTPTOMA 4651
Qy 1011 --PRKMTSTPELNPTSRIAEAMLOTTTRNQT-----PNSKL----- 1046
Db 4652 KSPKEMVQSPYEYKEAVLSGEDLDESTSPNMSLQDQELSLSSQPHGRIPHVPVTPSGK 4711
Qy 1047 -----VEVNPKS-----EDAG 1057
Db 4712 IYLYHYAEPPTGPFVEPPDLFFLTKTKAPVQGTATHVEKSPKEMVFRASPYKQAVLPAH 4771
Qy 1058 GAEGET--PHMLL-----RPHVEMPE-----VTPDMD 1082
Db 4772 EAQDESPPNMSLQDQELTMSSQPHGWIHPHNPHTGKIYLYHYAEPPTGPFVEPPDLF 4831
Qy 1083 YLPR-----VFNQ-----IINPM----- 1097
Db 4832 FLTKTKSPVQGTATQMAKSPPEMVSLLENKEAVLPAQDGLSEISIFPNMSLQPLDHE 4891
Qy 1098 --LSDE-----TNI-----CNGKPDGLTLRNGT 1120
Db 4892 LILSSQPHGWIPHHNTPGKIYLYHYAEPPTGPFVEPPDLFFLTKTKSPVQ-----SPT 4946
Qy 1121 LVA-----FRGH-----YFWMLSPF 1135
Db 4947 LIAKSPREMVSRSPYKQALLPGHGGKNSRAPNMSLQDQELTMSSQPGW----- 5001
Qy 1136 SPSPARRITEVWGPSPIDTTFVTRCNCEGKT-----FFKDSQ----- 1174
Db 5002 -----IPHPENT-----PGKIYLYHYAEPPTGPFVEPPDLFFLTKTKSKP 5040
Qy 1175 -YWRFTNDIKAGYKPIFKGFGGLTGQIVAAALSTAKYK----- 1212
Db 5041 VQW---TSAEIAKLPEIVS-----QTKHAKESVLRAPVEQHKNKSPSPNVS 5084
Qy 1213 -----NW-----PESVYFFKRGSGIQOYIY 1232
Db 5085 LQPLDQELSLFQPHGWIHPSPNTPGKIYLYHYAEPPTGPFVEPPDLFFLTK-----TT 5136

C;Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: kinase interaction domain homology <KIH>

Query Match 12.3%; Score 897.8; DB 2; Length 2897;
Best Local Similarity 14.8%; Pred. No. 1.2e-10;
Matches 464; Conservative 153; Mismatches 476; Indels 2052; Gaps 115;

```

QY 3 WKT-----LPIVLLLSLVFVIQVSSQELSCGRCFSEFREGRECDCA 47
Db 2 WPTRRRLVTKRGVDGPHFPLSLSLCL-----FGRGIECDIRI 39
QY 48 Q-----CKYDKCCDYESFCAVH-----NPT-----70
Db 40 QLPVVKQOHCK-----IEIHEQEAILHNFSTNPTQVNGSVIDEPVRLKHGD 86
QY 71 -----SPPSSKKAPPSGASQT 87
Db 87 VITIIDRSFRYENESLONGRKSTFPPRKIRQEPARRVSRSSFSDDPEKAQDSKAYSKI 146
QY 88 IKSTTKRSPK-----PPN-----KKKT-----KKVIESEETEEHSYSEN 122
Db 147 TEGKVSNGNELFDENLPNPLKRGAPTCKKSLVMTHTPPVLKLIK-----EQQPSGK 201
QY 123 QESS-----SSSSSSSSST-----137
Db 202 QESGSEIHVEYKAQSLVISPPAPSPRKTPVASDQRRRSCKTAPASSKSQTEVPKRGGER 261
QY 138 -----IWKIKSKNSAANRELOK-----KLKVDNK-----163
Db 262 VATCLOKRVSRISQSDILOMICSKRRSGASEANLIVAKGWADVVKLGAKQTQTKVIKHG 321
QY 164 ---KNRTKKKP-TPKPPV-----177
Db 322 PQRSMNKRRRRPATPKPKPVGVHVSQFSTGHANSPTIILIGKAHTEKVHVPARPVRLNNF 381
QY 178 -----VDEAG-----SGLDNGD-----189
Db 382 ISNQKMDFKEDLGSIAEMFKTPVKEQPLTSTCHIAISNENLLGKQFGQTDSEELLP 441
QY 190 -----FKVT-----TPD 196
Db 442 TSSEFGNVEFSAQNAKQSDKCSAPPLRQCIRENGNVAKTPRNTYKMTSLETXTSD 501
QY 197 TSTTOHNKVT---SPKITAKPINRP-----221
Db 502 TETPSKTVSTVNRSGRSTEFNIOKLPIVESKSEETNTEIVEICILKRGQXATILQORREG 561
QY 222 -----SLP-----224
Db 562 EMKEIERPPETKYKENIELKENDEKMKAMKASRTWGQKCAPMSDLTDLKSLPDTELMKDTA 621
QY 225 -----PNSDTSKETSLTVNK-----ETTVETK-----246
Db 622 RGQNLIQTQDHAKAPKSEKGIKTMPCOSLOPEPINTPTHKQOLKASLGKVGKBEALLA 691
QY 247 -----ETVTNKQTSDDGRE-----TKTSAKETOSTE 273
Db 682 VGKFRRTSGETHTHREPAGDGKSIREFKESPKQILDPAARVTGMKKWPTPKBEAQSLIE 741
QY 274 KTSAKDLATPSKVLAQTPKAET-----TTK-----299
Db 742 -----DLAGFKELFQTPGPSSESMDEKTKIACKSPPPESVDTPTSTKQWPKRSLRKAD 796
QY 300 -----GPAITTPKEP-----TPT-----TPK 315
Db 797 VBEEFILALRLKLTSPSAGKAMLTTPKAGDEKODIKAFMGTPVQKLDLAGTLPGSKRQLQTEK 856
QY 316 EPAST-----TPKEP-----TPTTIKSAPT-----335
Db 857 EKAQALEDLAGFKELFQTPGHTHEELVAAGTKIKIPCDSPQSDPVDPTSTKQRPKRSIRK 916
QY 336 -----TPKEPAP-----TTTAKSA 348

```

```

Db 917 ADVEGELLACRLNMPISAGKAMHTPK---PSVGEEDKIIIFVGTPTVQKLDLTENLTGSKRR 973
QY 349 PTTTPKEPAPT-----TTKEPAPTTTPKEPAPTTTPKEPAPT 383
Db 974 PQTPEEAQALEDLTGFKELFQTPGHTHEEVAAGTKTKMPCESSPESADT-----PIS 1027
QY 384 TKSAPTTTPKE-----PA 395
Db 1028 TRROFKPTPLEKRDYQKELSALKLTLTSGETTHDKVPGGEDKSINAFRETAQOKLDPA 1087
QY 396 PTTTPKAPTTTPKEPA-----PTTPK-----EPTPT-TPK 424
Db 1088 SVTQSKRHPKT-KEKAQPLEDLAGFKELFQTPVCTDRPTTHEKTKTIACKSQDPVDPTPT 1146
QY 425 EPAPTTKE-----PAPT-----436
Db 1147 SSKPQSKRSLRKVDVEEFPALRKRTSPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKLDLT 1206
QY 437 -----TPKEPA-----PTAPKK 448
Db 1207 ENLTGSKRRLQTPKEKAQALEDLAGFKELFQTRGHTBESMTNDKTAKVACKSSQPDLDKN 1266
QY 449 PA-----PTTPKEPAPT-----460
Db 1267 PASSKRRLKTSLSGKVGKVELLAVGKLTQTSGETTHHTHTPTGDKSKMAMPESPKOILD 1326
QY 461 -----TPKEPA-----PTTPKPSPTTPK-----EP--481
Db 1327 SAASLTGSKRQLRTPKGSVPEDLAGFIELFQTPSHTKE-SMTNEKTKTVSVASQPDLD 1385
QY 482 --APTTPKSAP-----490
Db 1386 VDTPTSSKPOPKRSLRKADTEEBFLAFRKQTPSAGKAMHTPKPAVGEEDKINTFLGTPVQ 1445
QY 491 -----TTTKEPA-----PT-----TTK-----SAPTT 507
Db 1446 KLDQPGNLPGSNRLQTRKEKAQALELTGFRFLFQTPCTDNPTADEKTTTKILCKSPQS 1505
QY 508 PKEPSPPTTKE-----PA-----520
Db 1506 DPADTPTNTKQPKRSLLKADVEEFLAFRLKTPSAGKAMHTPKAAVGEEDKINTFVGTP 1565
QY 521 -----PTTPKEPAPT-----TP-----KKP 535
Db 1566 VEKLDLLGNLPGSKRRPQTPKEKAKALEDLAGFKELFQTEGHTBESMTDDKI TEVSCKSP 1625
QY 536 APTTPKEPA-----544
Db 1626 QPDPVKTPTSSKQRLKLSLGKVGKVEEVLPGKLTQTSKTTOTTHRETAGDGKSIKAFKE 1685
QY 545 -----PTTPKEPA-----PTTTKPAPTAPK 565
Db 1686 SAKQMLDPANYGTGMERWPRTPKEEAQSLLEDLAGFKELFQTPDHTESITDDKTKIACK 1745
QY 566 EPAP---TTTPKET--APTTP-----KLTPTPT-----587
Db 1746 SPPPSMDPTSTTRRRPKTPLGKRDI VEELSALKQLTQTTHTDKVPGEDBKGINVPRETA 1805
QY 588 PEKLAB---TTPEKAPTTTP-----BELA-----PTTSEEPPTTPPEP 623
Db 1806 KQKLDPAASVTGSKRQPTPKQAQPLEDLAGFKELFQTPVCTDRPTTHEKTKTIACKRSP 1865
QY 624 AP-----TTPKAAA-----632
Db 1866 QPDPVGTPTIFKPQSKRSLRKADVEEESLALRKRTSPSVGKAMDTPKPAGDEKMKAFMG 1925
QY 633 -----PNTPKPEPA-----PTTPKEPAPTTPKE 654
Db 1926 TPVQKLDLPGNLPGSKRWPTPKKAQALEDLAGFKELFQTPGTDKPTTDEKTKTIACKS 1985
QY 655 PAP---TTTPKET-----APTTPK-----669
Db 1986 PQDPDPTPTASTKQRPKNLNRKADVEEFLALRKRTSPSAGKAMDTPKPAVSEKKNITFV 2045

```

670 QY -----GTAPTTLKBPAPTPPKAP-----KELAPTTTKEPTSTSDK----- 707
Db : : : : :
2046 ETPVQKLDLGNLPGSKRQ--POTPKKAEALBDLVGFKELFQTPGHTBESMTDDXITEV 2103
QY -----PAP-----TPPKGTAPT----- 719
Db : : : : :
2104 SCKSPQESFKTSRSKQRLKIPLVKMDKBEPLAVSKLRTSTGETTHTEFTGDSKSI 2163
QY -----TPK-----EPAPT-----TPKEPA----- 733
Db : : : : :
2164 KAFKESPKQLDPAASVTGSRRLRTKKEKARALEDLVDFKELFSAPGHTESMTIDKNT 2223
QY -----PTPKGTAPTTLKBPAPTPPKAPKELAP-----TTTKGPTSTTSKPA----- 778
Db : : : : :
2224 KIPCKSPPELTDTATSTKRCP--KTRPKVEKBELSAVERLLTQTSGOSTHKEPAGSDE 2282
QY -----PTTPKE-----TAPTT 789
Db : : : : :
2283 GIKVLKORAKKKNPVEEESRRRPRAPKKAQPLEDLAGFTELSETSGHTQESLTAGKA 2342
QY PKBPAPTPPKAPPTPETPTTTSVSTTPTTKBPTTIHKSPDESTPELSAETTP----- 845
Db TKIPCESPPELVVDTTASTKRLRLTRVQKVKEEPSAV--KFTQTSGETTADADKEPAGED 2401
QY -----KALENSKPEGVPTT-----KTPAA--TKPEMTTAK 875
Db : : : : :
2402 KGIKALKESAKQTPAPAAASVTGSRRLRPRAPRESAQALADLAGFKDPAAGHTESMT--D 2458
QY DKTERDLRTPE--TTAAPKMTKETAATTEKTE--SKITATTTQVTSITTQ-- 925
Db : : : : :
2459 DKTIKIPCKSPELEDTATSKRPRTRAQKVEKEELLAVGKLQTSGETHTDKEPVG 2518
QY -----DTPPKKLTTLKTTLAPKVTTKTITTEIMNKPEETAKPKORA----- 970
Db : : : : :
2519 EGKTKAFK-----QPAKRNVAEDVIGSRROPRAPKKAQPLEDLASQEL 2565
QY -----TNSKATPKPKETK-----APK----- 988
Db : : : : :
2566 SQTGHTTELANGAADSTAPKQTPDGSGLKISRVLRAKVPVGVVDVSTDRDPVKSQ 2625
QY -----KPTSTKPKTMRVVRKPKTTTPRKMTSTMP-----ELNP 1023
Db : : : : :
2626 SKSNTSLPPLPFKRGKGDSVGTGKRLCM-----PAPEIIEELPASKQORVAP 2676
QY -----IAPAMLQTTTR-----PNQTPNSKLVEVP-----KSEDAGGAGETPHMLLR 1069
Db : : : : :
2677 RARGKSSEPVVIMKRSRLTSAKRIEPAEELNSDMKTNKEEHLQDS----- 2723
QY PHVFMPEVTPDMVILPRVFNQGIILNPLSDETNICNGKPVVDGLTLNGLVAFRGHYF 1129
Db : : : : :
2724 -----VPE-----NKGISLRSRRODKT----- 2740
QY WMLSPFSPSPARRITEVW-----GIPSPIDTVFTRCNCEGT 1167
Db : : : : :
2741 -----EAEQOITEVFLAERTEINREKPKMTSPENDIQNP----- 2777
QY FFFKDSQYWRFTNDIKDAGYKPKPIKPGGGLTGQIIVAALSTAKYNWPESVFFKRGSSI 1227
Db : : : : :
2778 -----DDGARKPI----- 2785
QY QQYIYKQBPVQKCPGRRPALNVPVVGEMTQ-----VRRRRFERAIGPSQ 1271
Db : : : : :
2786 -----PRDKVTENKRLRSARQNESQPKVAEESGGQSAKVMQNKQKGGAGNSD 2837
QY THTR-----IQYSPARLAYODKGLVHNEVKVSIWGLPENVVTSALS--LPNIRKPDGDYD 1326
Db : : : : :
2838 SMCLRSRKTQSQAASTLESKSVOR-----VTRSVKRCAPENPKKAED----- 2879
QY VAFSKQDYNIDVPSRTARAITRS 1351
Db : : : : :
2880 -----NVCV-----KKITRS 2890

RESULT 14

25697
Hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-1229 <Full>
A:Cross-references: UNIPROT:Q94195; EMBL:U67956; PIDN:AA807691.1; GSPDB:GN00028; CESP:F16F9
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3
Query Match 12.2%; Score 890.1; DB 2; Length 1229;
Best Local Similarity 23.8%; Pred. No. 2.8e-11;
Matches 372; Conservative 116; Mismatches 395; Indels 679; Gaps 76;
QY 2 AWKTLPIYLLLLSVFVIQOVSSQELSCKG-RCFESPERGRECDCAQCKKYDKCCPDYE 60
Db 4 AW-VVSFAFLLGNVQSSLLSKTINSQSRDFKIVKHIKONTCTCSC---KCVPD-- 56
QY 61 SFCAEVHNTSPSPSKKAPPPSGASQTIKTTTKRSPKPNKKTKKVIKIEEITEHSVS 120
Db 57 -----APSN-----PFDVSTTSSIN-----N 73
QY 121 ENOESSSSSSSSSSSTIW-----KIKSSKNSAANRELOKLVKDNKNKRTKK 169
Db 74 DNDVIGSGSNTPTGSS-WFOEIEATVGGQTVKSEHNIDSSVEKKV----- 120
QY 170 KPTPKPVWDEAGSLDNGDFKVTTPDTSTTQ-----HNKVSTSPKI----- 211
Db 121 -----TTSDASITNAPTCKDSTTPEIITGIVVINSKESVT 158
QY 212 -----TTAKPINRPSLPNSDTSKETSIVNKEETVETKETTNTKOTSDGKEKT 263
Db 159 DMSTTSTSTLSP-----TTLELSPETLVSTDSSTST-EQTSPTDNTPEI 202
QY 264 TSAKETOSIEKTSADLAPTSKVLAKPTPRAETTTKGPALTTPKEPTTT--PKEPAST 320
Db 203 ASPMET-----NTTTE--ATTTSVEPSVSTLASEDETTVT 235
QY 321 TPKEPTPTTIKSAPTTPKEP-----APTITKSAPTTPKEPAPTTPKEPAPTTPK 377
Db 236 AIAESTTVIAEVSTTTEPTTAEITKSTT---KAPATTEPTTTEE--VTTTE 289
QY 378 EPAPTITKSAPTTPKEPAPTPP-----KKPAPTTPKEPAPTTPKEP--TPTTPKEPAPTKE 432
Db 290 AETSTTTSSETSTSK-----PTTPLIDNKIAGPATGK---PETTHFPVGTGTP--NFD 337
QY 433 PAPTPP-----KEPAPTAKKAPPTTPKEPAPTTPKEPAPTTPKE-----PSPTTPKEPAP 483
Db 338 TATETPFVAKSEDKMWLTKTAATETQOTTEVT-DGPEKEITKNVSIPIPTVPLVETT 396
QY 484 TTTKSAP-----TTTKEPAPT-----TTKSAPTTPKEPSPTTKE-----PAPT----- 522
Db 397 STTSTASKESDGEHTLLKLVTTADSDSTESATTVKPENEETTTKSHVVPKTKGTVKV 456
QY 523 TPK-----EPAPTTPKAPAP----- 537
Db 457 TPKLELSEDEPTEIT-KAPHEPKLLEKTYHFLVSDNFARVSEAKENDDYNHLDYNNHRE 515
QY 538 -----TTPKEPA-----PTTPKEPAPTTPKAPAPKEPA--PTTPKE 574
Db 516 AKBPITTEESSTTEETVTTTEPANTGNPTTEN---PTTTEQPTSTABSTTALPFTTEQ 572

Search completed: October 13, 2004, 12:00:49
Job time : 103.763 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 109.728 Seconds
(without alignment)
4449.477 Million cell updates/sec

Title: SEQ1-G
Perfect score: 7323
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGQTLKQVWYVNC 1361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | | | Description | |
|------------|--------|-------------|--------|----|-------------|-----------|
| Result No. | Score | Query Match | Length | ID | | |
| 1 | 7308.7 | 99.8 | 1404 | 2 | AAR26049 | MSF precu |
| 2 | 7308.7 | 99.8 | 1404 | 4 | AAB29773 | Human meg |
| 3 | 7308.7 | 99.8 | 1404 | 4 | AAB60568 | Human meg |
| 4 | 7308.7 | 99.8 | 1404 | 8 | ADM98014 | Human meg |
| 5 | 7305.7 | 99.8 | 1415 | 4 | AAU32262 | Novel hum |
| 6 | 7293.7 | 99.6 | 1404 | 7 | ADK58339 | Angiogene |
| 7 | 7043.9 | 96.2 | 1320 | 7 | ADK65819 | Angiogene |
| 8 | 7043.9 | 96.2 | 1311 | 8 | ADK67912 | Human ext |
| 9 | 6778.9 | 92.6 | 1270 | 8 | ADK67911 | Human ext |
| 10 | 6732.7 | 91.9 | 1299 | 4 | AAW24322 | Human EST |
| 11 | 3552 | 48.5 | 902 | 4 | AAB29778 | Human MSF |
| 12 | 2929 | 40.0 | 551 | 4 | ABUS3253 | Human tes |
| 13 | 2920 | 39.9 | 546 | 4 | ABUS3252 | Human tes |
| 14 | 2850.9 | 38.9 | 538 | 5 | AAO18834 | 3' cartil |
| 15 | 2757 | 37.6 | 513 | 4 | ABUS3254 | Human tes |
| 16 | 2232.9 | 30.5 | 452 | 2 | AAR80041 | Human meg |
| 17 | 2004 | 27.4 | 472 | 4 | AAB60569 | Bovine MS |
| 18 | 1566.9 | 21.4 | 232 | 5 | AAU11261 | Human HAP |
| 19 | 1413.6 | 19.3 | 5179 | 4 | AAW24516 | C899P pre |
| 20 | 1413.6 | 19.3 | 5179 | 6 | ABP55365 | Human col |
| 21 | 1413.6 | 19.3 | 5179 | 6 | ABO07258 | Human p53 |
| 22 | 1413.6 | 19.3 | 5179 | 7 | ADDA48091 | Human Pro |
| 23 | 1413.6 | 19.3 | 5179 | 7 | ADDA4998 | Human Pro |
| 24 | 1189.7 | 16.2 | 8991 | 6 | ABU08487 | S. pneumo |
| 25 | 1179 | 16.1 | 1664 | 2 | AAW43106 | C. thermo |

| ALIGNMENTS | | | | | | | | | |
|-----------------------------|---|---------------------|--|--|--|--|--|--|--|
| RESULT 1 | | | | | | | | | |
| AAR26049 | | | | | | | | | |
| ID | AAR26049 standard; protein; 1404 AA. | | | | | | | | |
| XX | AC | AAR26049; | | | | | | | |
| XX | XX | | | | | | | | |
| DT | 25-MAR-2003 (revised) | | | | | | | | |
| DT | 02-FEB-1993 (first entry) | | | | | | | | |
| XX | XX | | | | | | | | |
| DE | MSF precursor. | | | | | | | | |
| XX | XX | | | | | | | | |
| KW | Megakaryocyte colony stimulating factor; secretion signal; meg-CSF; | | | | | | | | |
| KW | stability; proteolytic cleavage; adhesion; alternative splicing. | | | | | | | | |
| XX | OS | Synthetic. | | | | | | | |
| XX | XX | | | | | | | | |
| PH | Key | Location/Qualifiers | | | | | | | |
| FT | Region | 1..26 | | | | | | | |
| FT | | /label= Exon_I | | | | | | | |
| FT | Region | 26..67 | | | | | | | |
| FT | | /label= Exon_II | | | | | | | |
| FT | Region | 67..107 | | | | | | | |
| FT | | /label= Exon_III | | | | | | | |
| FT | Region | 107..157 | | | | | | | |
| FT | | /label= Exon_IV | | | | | | | |
| FT | Region | 157..200 | | | | | | | |
| FT | | /label= Exon_V | | | | | | | |
| FT | Region | 200..1141 | | | | | | | |
| FT | | /label= Exon_VI | | | | | | | |
| FT | Region | 1166..1212 | | | | | | | |
| FT | | /label= Exon_VIII | | | | | | | |
| FT | Region | 1213..1266 | | | | | | | |
| FT | | /label= Exon_IX | | | | | | | |
| FT | Region | 1266..1331 | | | | | | | |
| FT | | /label= Exon_X | | | | | | | |
| FT | Region | 1331..1373 | | | | | | | |
| FT | | /label= Exon_XI | | | | | | | |
| FT | Region | 1373..1404 | | | | | | | |
| FT | | /label= Exon_XII | | | | | | | |
| FT | Region | 1411..1166 | | | | | | | |
| FT | | /label= Exon_VII | | | | | | | |
| WO9213075-A1. | | | | | | | | | |
| 06-AUG-1992. | | | | | | | | | |
| 17-JAN-1992; 92WO-US000433. | | | | | | | | | |

Adl23265 Human MUC
Abb69806 Drosophil
Abp43908 MUC5B par
Abus3144 Human tes
Abus3255 Human tes
Aag38942 Arabidops
Abb71434 Drosophil
Abb61364 Drosophil
Abb58595 Drosophil
Aao18833 5' cartil
Aab68397 Drosophil
Abb60403 Drosophil
Abu53141 Human tes
Adi39110 Cancer/an
Adi21202 Novel hum
Abu53143 Human tes
Adi21666 Novel hum
Abu53155 Human tes
Abu53150 Human tes
Abu53149 Human tes

18-JAN-1991; 91US-00643502.
10-SEP-1991; 91US-00757022.

(GEMV) GENETICS INST INC.

Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG;
WPI; 1992-284660/34.
N-P5DB; AAQ27223.

New human mega-karyocyte stimulating factors - for treating immune deficiencies, cancer, exposure to radiation or drugs, bacterial and viral infections, etc.

Claim 1, 2 and 3; Fig 1; 87pp; English.

The sequence given is a full length translation from the megakaryocyte stimulating factor (MSF) precursor. The sequence covered by exons II, III and IV encodes megakaryocyte stimulating factor (MSF). This sequence is modified by the addition of an N-terminal sequence encoding a secretory leader, an initiating methionine proceeding exon II and a terminating codon following exon IV. The cDNA sequence given contains sequences derived from human megakaryocyte colony stimulating factor (meg-CSF). Exon I contains the initiating methionine, and encodes a classical mammalian protein secretion signal sequence. The sequence encoding the original meg-CSF includes exons II-IV and is thought to terminate in the region between amino acid residues 134 - 147. The primary transcript of this gene may be cleaved in different ways to yield a family of mRNA's each encoding a different MSF protein. Exons V and VI are thought to be related to the activity of the factor and are also implicated in the stability, folding and processing of the molecule. These exons are also thought to play a role in the observed synergy of MSF with other cytokines. Exons V - XII are believed to be implicated in the processing or folding of the appropriate structure of the resulting factor, i.e. one or more of these exons may contain sequences which direct proteolytic cleavage, adhesion, organisation of the cellular matrix or extracellular matrix processing. Both naturally occurring and non-naturally occurring MSF's may be characterised by various combinations of alternatively spliced exons from this sequence, with the exons spliced together in differing orders to form different members of the MSF family. (Updated on 25-MAR-2003 to correct PN field.)

XX
SQ

Sequence 1404 AA;

Query Match 99.8%; Score 7308.7; DB 2; Length 1404;
Best Local Similarity 96.9%; Pred. No. 1.8e-176;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTIPIYLILLLSVFVTIQOVSSODLSSCAGCGSGYGRDATCNCDYNCOHMECCPDF 60
DB 1 MAWKTIPIYLILLLSVFVTIQOVSSODLSSCAGCGSGYGRDATCNCDYNCOHMECCPDF 60
QY 61 KRVTCAELSCGRCFESFERGREDCDQAOCCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRVTCAELSCGRCFESFERGREDCDQAOCCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPASGQTIKSTTKRSPKPNKKTKKVIESEITE----- 156
DB 121 PPASGQTIKSTTKRSPKPNKKTKKVIESEITEHSVSNOESSSSSSSSSSTIW 180
QY 157 -----VKDNKNRTKKKTPKPPVVDEAGSLDNGDFKVTTEDTST 197
DB 181 KIKSSKNSAANRELQKLKYVDKNKNETKKKPKPPVVDEAGSLDNGDFKVTTEDTST 240
QY 198 TOHNVKSTSFKITTAQINPRPSLPNNSDTSKETSLTNVNETTVEKETTTNKQSTDG 257
DB 241 TOHNVKSTSFKITTAQINPRPSLPNNSDTSKETSLTNVNETTVEKETTTNKQSTDG 300
QY 258 KEKTTSAKETOSIEKTSADKLAPTSKVLAQTPKAETTGTGPALTTPKEBPTTPPKBPAS 317
DB

RESULT 2

QY 978 PQKPTKAPKPTSTKKPTMPVRVKPTTPTPKMTSTMPELNPTSRIAEAMLQTTTPN 1037
DB 1021 PQKPTKAPKPTSTKKPTMPVRVKPTTPTPKMTSTMPELNPTSRIAEAMLQTTTPN 1080
QY 1038 QTPNSKLVEVNPKEDEAGAGETPHMLLRPHVFMPEVTPMDYLPVRVNOGIIINPMLS 1097
DB 1081 QTPNSKLVEVNPKEDEAGAGETPHMLLRPHVFMPEVTPMDYLPVRVNOGIIINPMLS 1140
QY 1098 DETNLCNGKPDVGLTTLNGLTVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFT 1157
DB 1141 DETNLCNGKPDVGLTTLNGLTVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFT 1200
QY 1158 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQIIVAAALSTAKYKNWPSVY 1217
DB 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQIIVAAALSTAKYKNWPSVY 1260
QY 1218 FFKGGGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRPERAIGPSQTHIRIQY 1277
DB 1261 FFKGGGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRPERAIGPSQTHIRIQY 1320
QY 1278 SPARLAYODKGLHNEVKVSIILWGLPNVNTSAISLENIKRGDGYDYAFKSDQYINIDV 1337
DB 1321 SPARLAYODKGLHNEVKVSIILWGLPNVNTSAISLENIKRGDGYDYAFKSDQYINIDV 1380
QY 1338 PSRTARAITTRSGOTLSKVVWVNC 1361
DB 1381 PSRTARAITTRSGOTLSKVVWVNC 1404

RESULT 4
ADM98014
ID ADM98014 standard; protein; 1404 AA.
AC ADM98014;
XX
XX 01-JUL-2004 (first entry)
XX Human megakaryocyte stimulating factor (MSF).
XX lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
XX CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
XX megakaryocyte stimulating factor; MSF.
XX
XX Homo sapiens.
XX
XX US2004072741-A1.
XX 15-APR-2004.
XX
XX 02-JUL-2001; 2001US-00897188.
XX
XX 23-APR-1999; 99US-00298970.
XX 24-APR-2000; 2000US-00556246.
XX
XX (JAYG/) JAY G D.
XX
XX Jay GD;
XX
XX WPI; 2004-373948/35.
XX N-PSDB; ADM98015.
XX
XX New tribonectin polypeptides and polynucleotides for lubricating joints
XX or other tissues to prevent or treat Camptodactyl-arthropathy -
XX pericarditis syndrome or osteoarthritis.
XX
XX Claim 1; SEQ ID NO 1; 34pp; English.
XX
XX The invention relates to a lubricating polypeptide and at least one O-
XX linked oligosaccharide. The composition and methods are useful for
XX lubricating joints or other tissues to prevent or treat camptodactyl-
XX arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
XX sequence represents the amino acid sequence of the human megakaryocyte

CC stimulating factor (MSF).
XX
SQ Sequence 1404 AA;
Query Match 99.8%; Score 7308.7; DB 8; Length 1404;
Best Local Similarity 96.9%; Pred. No. 1.8e-176;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCHYMCCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCHYMCCPDF 60
QY 61 KRVTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYSPCAEVNPTSPPSSKKAP 120
DB 61 KRVTAELSCKGRCFESFERGECDCDAQCKKYDKCCPDYSPCAEVNPTSPPSSKKAP 120
QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEETI----- 156
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEETI----- 180
QY 157 -----VKDNKNRTKKKPTPKPVVDGAGSLDNGDFKVTTPDTST 197
DB 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPVVDGAGSLDNGDFKVTTPDTST 240
QY 198 TOHNKVSTSPKITTAKPINRPSLPNSDSKETSLSLVNKETTVETKETTNNKQTSDDG 257
DB 241 TOHNKVSTSPKITTAKPINRPSLPNSDSKETSLSLVNKETTVETKETTNNKQTSDDG 300
QY 258 KEKTTSAKETQSIBKTSAKOLAPTISKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 317
DB 301 KEKTTSAKETQSIBKTSAKOLAPTISKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360
QY 318 TTPKEPTTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 377
DB 361 TTPKEPTTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 420
QY 378 APITTKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 437
DB 421 APITTKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 480
QY 438 EPAPTAPKPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 497
DB 481 EPAPTAPKPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 540
QY 498 TTKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 557
DB 541 TTKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 600
QY 558 APAPTAPKPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 617
DB 601 APAPTAPKPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 660
QY 618 PEPAPATTPKAAAPNTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 677
DB 661 PEPAPATTPKAAAPNTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 720
QY 678 APPTPKKPAKELAPTTTKPTSTSDKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 737
DB 721 APPTPKKPAKELAPTTTKPTSTSDKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 780
QY 738 TAPTTLKEAPATTPKPAKELAPTTTKGPTSTSDKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 797
DB 781 TAPTTLKEAPATTPKPAKELAPTTTKGPTSTSDKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPK 840
QY 798 KPAPTTPPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSABPTPKALENSKEGCVPT 857
DB 841 KPAPTTPPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSABPTPKALENSKEGCVPT 900
QY 858 TKTPAAKPEMTTAKDKTERDLRTPPTTTAAAPKMTKETATTTTEKTTESKITATTTQV 917
DB 901 TKTPAAKPEMTTAKDKTERDLRTPPTTTAAAPKMTKETATTTTEKTTESKITATTTQV 960
QY 918 TSTTTQDTTFFKLTLLKTLTLAPKVTTTKKLTITTTTSMNKPEETAKPKORATNSKATTPK 977

Db 961 TSTTTQDTDFKUTLLTKTTTLLAPKVTTKTKITTTTEINNKPETAKKDRATNSKATPK 1020
QY 978 PQKPTKAPKPTSTKKPKTMRVRKPKTTTTPRKMTSTMPELNPTSRIABAMLQTTTRPN 1037
Db 1021 PQKPTKAPKPTSTKKPKTMRVRKPKTTTTPRKMTSTMPELNPTSRIABAMLQTTTRPN 1080
QY 1038 QTPNSKLVFNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQIINPMLS 1097
Db 1081 QTPNSKLVFNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQIINPMLS 1140
QY 1098 DETNIONGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITTEWGLPSPIDTVFT 1157
Db 1141 DETNIONGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITTEWGLPSPIDTVFT 1200
QY 1158 RCNCEGKTTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIIVAALSTAKYKNWPSVY 1217
Db 1201 RCNCEGKTTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIIVAALSTAKYKNWPSVY 1260
QY 1218 FFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERAIGPSQTHIRIQY 1277
Db 1261 FFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERAIGPSQTHIRIQY 1320
QY 1278 SPARLAYQDKGLHNEVKVSLWRGLPNVWTSALSLEPNIRKPDGYDYAFSKOQYXNIDV 1337
Db 1321 SPARLAYQDKGLHNEVKVSLWRGLPNVWTSALSLEPNIRKPDGYDYAFSKOQYXNIDV 1380
QY 1338 PSRTARAITRSGQTLKSVWYNCP 1361
Db 1381 PSRTARAITRSGQTLKSVWYNCP 1404
RESULT 5
AAU32262
ID AAU32262 standard; protein; 1415 AA.
AC AAU32262;
XX
DT 18-DEC-2001 (first entry)
XX Novel human secreted protein #2753.
DE Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX Homo sapiens.
OS
XX WO200179449-A2.
PN
XX 25-OCT-2001.
PD
XX 16-APR-2001; 2001WO-US008656.
PF
XX 18-APR-2000; 2000US-00552929.
PR
XX 26-JAN-2001; 2001US-00770160.
PR
XX (HYSE-) HYSEQ INC.
FA
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-611725/70.
DR
XX Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy.
PT
XX Claim 20; Page 573; 765pp; English.
PS
XX The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
XX Sequence 1415 AA;
QY Query Match 99.8%; Score 7305.7; DB 4; Length 1415;
Best Local Similarity 96.9%; Pred. No. 2.1e-176;
Matches 1360; Conservative 1; Mismatches 0; Indels 43; Gaps 1;
1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
2 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 71
61 KRVCCTAELSCGRCFCSEFREGRECDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
72 KRVCCTAELSCGRCFCSEFREGRECDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 131
121 PPSGASQTIKSTTKRSPKPNKTKKVISEEITE----- 156
132 PPSGASQTIKSTTKRSPKPNKTKKVISEEITEEHSVSENESSSSSSSSSTIM 191
157 -----VKDNKNRTKKKTPKPVVDAGSLDNGDFKVTTPDTST 197
192 KIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPVVDAGSLDNGDFKVTTPDTST 251
198 TQHNKVSSTPKITTAKPINRPSLPNSDTSKETSILVNKETTVEKTTTTNKQTSIDG 257
252 TQHNKVSSTPKITTAKPINRPSLPNSDTSKETSILVNKETTVEKTTTTNKQTSIDG 311
258 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTKGPAITTPKEPTTPPKBPAS 317
312 KEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTKGPAITTPKEPTTPPKBPAS 371
318 TTPKEPTPTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTK 377
372 TTPKEPTPTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTK 431
378 APTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 437
432 APTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 491
438 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 497
492 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 551
498 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 557
552 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 611
558 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 617
612 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 671
618 PEEPAPTTPKAAANPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 677
672 PEEPAPTTPKAAANPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 731
678 APTTPKPAKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 737
732 APTTPKPAKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 791
738 TAPTTLKBPAPTTTPKPAKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 797

| | | | |
|----|------|--|------|
| Qy | 541 | PAPTPKEPAPPTTKKPAFTAPKEPAPPTPKETAPPTPKLPTPTTPEKLPAPTTPKEPAPT | 600 |
| Db | 500 | PAPTPKEPAPPTTKKPAFTAPKEPAPPTPKETAPPTPKLPTPTTPEKLPAPTTPKEPAPT | 559 |
| Qy | 601 | TPRELAPPTPEPPTTPEEPAPPTPKAAAPNTPKEPAPPTPKAPPTTPKEPAPPTPK | 660 |
| Db | 560 | TPRELAPPTPEPPTTPEEPAPPTPKAAAPNTPKEPAPPTPKAPPTTPKEPAPPTPK | 619 |
| Qy | 661 | TAPTPKGTATTLKEPAPPTPKKAPKELAPPTTKETPTSTSDAPAPPTPKGAPPTPK | 720 |
| Db | 620 | TAPTPKGTATTLKEPAPPTPKKAPKELAPPTTKETPTSTSDAPAPPTPKGAPPTPK | 679 |
| Qy | 721 | EPAPTPKEPAPPTPKGAPPTTLKEPAPPTPKKAPKELAPPTTKGPTSTTSDKAPPTP | 780 |
| Db | 680 | EPAPTPKEPAPPTPKGAPPTTLKEPAPPTPKKAPKELAPPTTKGPTSTTSDKAPPTP | 739 |
| Qy | 781 | KETAPTPKEPAPPTPKKAPPTTPEPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE | 840 |
| Db | 740 | KETAPTPKEPAPPTPKKAPPTTPEPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE | 799 |
| Qy | 841 | PTPKALENSPKPEGVPTTKTAATKPEMTTTAKDKTTTERDLRTTPEITTAAPKTKETAT | 900 |
| Db | 800 | PTPKALENSPKPEGVPTTKTAATKPEMTTTAKDKTTTERDLRTTPEITTAAPKTKETAT | 859 |
| Qy | 901 | TTEKTTESKIDATTTQVTSITTTQDTPPKITTLTKTTLAPKVTTKTITTEIMNKPEE | 960 |
| Db | 860 | TTEKTTESKIDATTTQVTSITTTQDTPPKITTLTKTTLAPKVTTKTITTEIMNKPEE | 919 |
| Qy | 961 | TAKPKDRATNSKATTPKPKQPKAPKPTSTKPKTMPRVKRPKTTTPPRKWTSTMPBLN | 1020 |
| Db | 920 | TAKPKDRATNSKATTPKPKQPKAPKPTSTKPKTMPRVKRPKTTTPPRKWTSTMPBLN | 979 |
| Qy | 1021 | PTSRIAEAMLOTTTRPQNTNSKLVEVNPKSEDAGGACGETPHMLLRPHVFWPEVTPDMD | 1080 |
| Db | 980 | PTSRIAEAMLOTTTRPQNTNSKLVEVNPKSEDAGGACGETPHMLLRPHVFWPEVTPDMD | 1039 |
| Qy | 1081 | YLPRVFNQGIILNPLSDETNICMGKPDVGLTTLRNGTLVAFRGHYFWMLSPEFSPSPAR | 1140 |
| Db | 1040 | YLPRVFNQGIILNPLSDETNICMGKPDVGLTTLRNGTLVAFRGHYFWMLSPEFSPSPAR | 1099 |
| Qy | 1141 | RITEWGIPIGPDITVTRFNCCEGTFPFKDSQYWRFTNDIKDAGYKPIPKFGFGLGQI | 1200 |
| Db | 1100 | RITEWGIPIGPDITVTRFNCCEGTFPFKDSQYWRFTNDIKDAGYKPIPKFGFGLGQI | 1159 |
| Qy | 1201 | VAALSTAKYKNWESVYFVRKGGSIQYIYKQEPVQKCPGRPALNTPVIGMTQVRRR | 1260 |
| Db | 1160 | VAALSTAKYKNWESVYFVRKGGSIQYIYKQEPVQKCPGRPALNTPVIGMTQVRRR | 1219 |
| Qy | 1261 | FERAIGPSQTHIRIQSPARLAYQDKGVHLNHEVKVSIWRLGNVVTSAISLBNIRKPD | 1320 |
| Db | 1220 | FERAIGPSQTHIRIQSPARLAYQDKGVHLNHEVKVSIWRLGNVVTSAISLBNIRKPD | 1279 |
| Qy | 1321 | GVDYAFSKQYQYINIDVPSTARAITTRSGOTLSKWYNCP | 1361 |
| Db | 1280 | GVDYAFSKQYQYINIDVPSTARAITTRSGOTLSKWYNCP | 1320 |

RESULT 8
ADK67912
ID ADK67912 standard; protein; 1311 AA.
XX ADK67912;
AC
XX
DT 06-MAY-2004 (first entry)
XX Human extracellular messenger (EXMES) polypeptide.
DE Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;
KW antisthmatic; antinflammatory; antidiabetic; neuroprotective;
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic;
KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
KW cytostatic; gene therapy.

| | | |
|----|----|--|
| XX | OS | Homo sapiens. |
| XX | PH | Key |
| FT | FT | Peptide |
| FT | FT | Location/Qualifiers |
| FT | FT | 1..30 |
| FT | FT | /label= Signal peptide |
| FT | FT | /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30 |
| XX | XX | according to identification method" |
| XX | XX | WO2004013292-A2. |
| XX | XX | 12-FEB-2004. |
| XX | XX | 30-JUL-2003; 2003WO-US024084. |
| XX | XX | 02-AUG-2002; 2002US-0400810P. |
| XX | XX | 19-SEP-2002; 2002US-0412197P. |
| XX | XX | 04-OCT-2002; 2002US-0416004P. |
| XX | XX | 08-NOV-2002; 2002US-0424862P. |
| XX | XX | (INCY-) INCYTE CORP. |
| XX | XX | Elliot VS, Khare R, Tran UK, Swarnakar A, Marquis JP; Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J; Lee SY; |
| XX | XX | WPI; 2004-157116/15. |
| XX | XX | N-PSDB; ADK67917. |
| XX | XX | New extracellular messengers and nucleic acids, useful for diagnosing, treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or autoimmune thyroiditis. |
| XX | XX | Claim 60; SEQ ID NO 5; 165pp; English. |
| XX | XX | The present sequence is that of novel human extracellular messenger (EXMES) Incyte ID NO: 751018CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, fungal, osteoporosis, pancreatitis, psoriasis, viral, bacterial, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or their encoding polynucleotides for facilitating the drug discovery process, including determining of efficacy, dosage, toxicity and pharmacology, and for investigating the pathogenesis of diseases and medical conditions. |
| XX | XX | Sequence 1311 AA; |

| | | | |
|----|-----|---|-----|
| Qy | 1 | MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF | 60 |
| Db | 1 | MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF | 60 |
| Qy | 61 | KRYCTAELSCKGRCFESFERGECDDAOCKYVDKCCPDYSESCAEVHNPTSPSSKAP | 120 |
| Db | 61 | KRYCTAELSCKGRCFESFERGECDDAOCKYVDKCCPDYSESCAEVHNPTSPSSKAP | 105 |
| Qy | 121 | PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEITEVKNKNKNTKKKPTPKPPVDEAG | 180 |
| Db | 106 | -----EVDKNKNKNTKKKPTPKPPVDEAG | 130 |

Query Match 96.2%; Score 7043; DB 8; Length 1311;
Best Local Similarity 96.3%; Pred. No. 8.4e-170;
Matches 1310; Conservative 0; Mismatches 1; Indels 50; Gaps 1;

QY 181 SGLDNGDFVTTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLTNKETT 240
DB 131 SGLDNGDFVTTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLTNKETT 190
QY 241 VETKETTNTKQSTGDKETTSKAKTSIEKTSKADLAPTSKVLAKPTPKAETTTKQPA 300
DB 191 VETKETTNTKQSTGDKETTSKAKTSIEKTSKADLAPTSKVLAKPTPKAETTTKQPA 250
QY 301 LTTPEKPTPTPKPEASTTKEPTPTTIKSAPTTPKGPAPTPTTKSAPTTKPEAPTTTKE 360
DB 251 LTTPEKPTPTPKPEASTTKEPTPTTIKSAPTTPKGPAPTPTTKSAPTTKPEAPTTTKE 310
QY 361 PAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 420
DB 311 PAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 370
QY 421 PKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 480
DB 371 PKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 430
QY 481 PTTTTSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 540
DB 431 PTTTTSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 490
QY 541 PAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 600
DB 491 PAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 550
QY 601 TPESLAPTTPPEPTPTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPEEP 660
DB 551 TPESLAPTTPPEPTPTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPEEPAPTTTPEEP 610
QY 661 TAPTTPKGTATTLKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 720
DB 611 TAPTTPKGTATTLKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 670
QY 721 EPAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 780
DB 671 EPAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 730
QY 781 KETAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 840
DB 731 KETAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKE 790
QY 841 PTPKALENSKPEGVPTTKTAAKPEMTTAKDKTTERDLRTPETTTAAKPKTKETAT 900
DB 791 PTPKALENSKPEGVPTTKTAAKPEMTTAKDKTTERDLRTPETTTAAKPKTKETAT 850
QY 901 TTEKTTESKITATTTQVTSSTTQDTTPFKITLTKTTTLAPKVTTKKTIITTEIMNKPEE 960
DB 851 TTEKTTESKITATTTQVTSSTTQDTTPFKITLTKTTTLAPKVTTKKTIITTEIMNKPEE 910
QY 961 TAKPKDRATNSKATTPKQKPTKAPKKTSTKPKKTMPRVRKPKTTPPRKMTSTMPBLN 1020
DB 911 TAKPKDRATNSKATTPKQKPTKAPKKTSTKPKKTMPRVRKPKTTPPRKMTSTMPBLN 970
QY 1021 PTSRTAEAMLQTTTRPNQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1080
DB 971 PTSRTAEAMLQTTTRPNQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1030
QY 1081 YLPRVFNQGIINPMLSDETNICNGKVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPAR 1140
DB 1031 YLPRVFNQGIINPMLSDETNICNGKVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPAR 1090
QY 1141 RITEVWGIPTSIDTFTTRCNCEGKTFPFKDSQYWRFTNDIKDAGYFKPIFKGFGGLTQOI 1200
DB 1091 RITEVWGIPTSIDTFTTRCNCEGKTFPFKDSQYWRFTNDIKDAGYFKPIFKGFGGLTQOI 1150
QY 1201 VAAALSTAKYKNWPESVYFKRGSGTQOYIYKQEPVQKCPGRPALNYPVYGMTQVRRRR 1260
DB 1151 VAAALSTAKYKNWPESVYFKRGSGTQOYIYKQEPVQKCPGRPALNYPVYGMTQVRRRR 1210
QY 1261 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEKVSILWRGLPNVWVTSAISLENIRKPD 1320

DB 1211 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEKVSILWRGLPNVWVTSAISLENIRKPD 1270
QY 1321 GYDYIAFASKQOYNNIDVPSRTARAITTRSGQTLSKWYNCP 1361
DB 1271 GYDYIAFASKQOYNNIDVPSRTARAITTRSGQTLSKWYNCP 1311
RESULT 9
ADK67911
ID ADK67911 standard; protein; 1270 AA.
XX AC ADK67911;
XX DT 06-MAY-2004 (first entry)
XX Human extracellular messenger (EXMES) polypeptide.
DE Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;
XX anti-infective; anti-inflammatory; antidiabetic; neuroprotective;
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic;
KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
KW cytostatic; gene therapy.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..29
FT /label= Signal peptide
FT /note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29
XX according to identification method"
WO2004013292-A2.
PD 12-FEB-2004.
XX 30-JUL-2003; 2003WO-US024084.
XX 02-AUG-2002; 2002US-0400810P.
PR 19-SEP-2002; 2002US-0412197P.
PR 04-OCT-2002; 2002US-0416004P.
PR 08-NOV-2002; 2002US-0424862P.
XX (INCY-) INCYTE CORP.
XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;
PI Richardson TW, Emerling BW, Lindquist EA, Chawia NK, Ramkumar J;
PI Lee SY;
XX WPI; 2004-1571116/15.
DR N-PSDB; ADK67916.
XX New extracellular messengers and nucleic acids, useful for diagnosing,
PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
PT autoimmune thyroiditis.
XX Claim 59; SEQ ID NO 4; 165pp; English.
XX The present sequence is that of novel human extracellular messenger
CC (EXMES) Incyte ID NO: 7513017CD1 polypeptide. The protein shows homology
CC to human megakaryocyte stimulating factor. The invention provides EXMES
CC polynucleotides and polypeptides, as well as expression vectors, host
CC cells, antibodies, agonists and antagonists, and methods for diagnosing,
CC creating or preventing disorders associated with aberrant expression of
CC EXMES, especially autoimmune and inflammatory disorders, cell
CC proliferative disorders and endocrine disorders, e.g. adult respiratory
CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's
CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,
CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,
CC parasitic, protozoal or helminthic infections, cancers, autoimmune
CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.
CC Embodiments also provide methods for using the purified EXMES and/or

CC their encoding polynucleotides for facilitating the drug discovery
CC process, including determining of efficacy, dosage, toxicity and
CC pharmacology, and for investigating the pathogenesis of diseases and
XX medical conditions.
XX

SQ Sequence 1270 AA;

Query Match 92.6%; Score 6778.9; DB 8; Length 1270;
Best Local Similarity 93.2%; Pred No. 3.9e-163;
Matches 1269; Conservative 0; Mismatches 1; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRGEGYSDATCNDYNCQHMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
QY 61 KRVCIAELSCGRCEFERECDCDAQCKYDKCCDPDYSEFCAEVHNPTSPSSKKAP 120
DB 26 -----ELSCGRCEFERECDCDAQCKYDKCCDPDYSEFCA----- 64
QY 121 PPASGASQTIKSTTKRSPKPPNKKTKKVIKVEEIEIVKDNKKNRKTKKPTPKPPVDEAG 180
DB 65 -----EVDNKKNRKTKKPTPKPPVDEAG 89
QY 181 SGLDNGDFKVTPTDSTQHNVKSTSPKITTAKPINRPSLPNSDTSKETSIAVKNKETT 240
DB 90 SGLDNGDFKVTPTDSTQHNVKSTSPKITTAKPINRPSLPNSDTSKETSIAVKNKETT 149
QY 241 VETKETTITNKQSTDKGKETSIAKETQSIKTSIAKOLAPTSKVLAKPTPKAETTTKGA 300
DB 150 VETKETTITNKQSTDKGKETSIAKETQSIKTSIAKOLAPTSKVLAKPTPKAETTTKGA 209
QY 301 LTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKEPAPTITKE 360
DB 210 LTTPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKEPAPTITKE 269
QY 361 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKPEAPTTTPKPEAPTTTPKEPTPTT 420
DB 270 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKPEAPTTTPKPEAPTTTPKEPTPTT 329
QY 421 PKPEAPTTTPKEPAPTTPKPEAPTTAPKPEAPTTTPKPEAPTTTPKEPAPTTPKEPA 480
DB 330 PKPEAPTTTPKEPAPTTPKPEAPTTAPKPEAPTTTPKPEAPTTTPKEPAPTTPKEPA 389
QY 481 PTTTKSAPTTPKPEAPTTTKSAPTTPKEPSPTTTKEPAPTTPKPEAPTTTPKPEAPTTPKE 540
DB 390 PTTTKSAPTTPKPEAPTTTKSAPTTPKEPSPTTTKEPAPTTPKPEAPTTTPKPEAPTTPKE 449
QY 541 PAPTTPKEPAPTTPKPEAPTTAPKPEAPTTTPKETAAPTTPKLTPTTTPKLAAPTTPKEKAPT 600
DB 450 PAPTTPKEPAPTTPKPEAPTTAPKPEAPTTTPKETAAPTTPKLTPTTTPKLAAPTTPKEKAPT 509
QY 601 TPTELAPTTPEEPTPTTPPEPAPTTPKAAANPTPKPEAPTTTPKPEAPTTTPKEPAPTTPKE 660
DB 510 TPTELAPTTPEEPTPTTPPEPAPTTPKAAANPTPKPEAPTTTPKPEAPTTTPKEPAPTTPKE 569
QY 661 TAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTTPKGTAPTTPK 720
DB 570 TAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTTPKGTAPTTPK 629
QY 721 EPAPTTPKPEAPTTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTTP 780
DB 630 EPAPTTPKPEAPTTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKPEAPTTTP 689
QY 781 KETAPTTPKPEAPTTTPKPKAPKELAPTTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 840
DB 690 KETAPTTPKPEAPTTTPKPKAPKELAPTTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPELSAE 749
QY 841 PTPKALENSKPEGVPTTKTAAKTPKEMTTTAKDKTTERDLRTTPEPTTAAPKMTKETAT 900
DB 750 PTPKALENSKPEGVPTTKTAAKTPKEMTTTAKDKTTERDLRTTPEPTTAAPKMTKETAT 809
QY 901 TTBKTTESKITATTTQVTSSTTTQDTPPKITLTKTTTLAPKVTITTKKTTITTEIMNKPEE 960
|||||

DB 810 TTEKTTESKITATTTQVTSSTTTQDTPPKITLTKTTTLAPKVTITTKKTTITTEIMNKPEE 869
QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMTPRVRKPKTTPTPRKMSTWPELN 1020
DB 870 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMTPRVRKPKTTPTPRKMSTWPELN 929
QY 1021 PTSIAEAMLOTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPMD 1080
DB 930 PTSIAEAMLOTTTRPNQTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPMD 989
QY 1081 YLPRVFNQGIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPESSPAR 1140
DB 990 YLPRVFNQGIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPESSPAR 1049
QY 1141 RITEVWGIPIIDVTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGSLTQOI 1200
DB 1050 RITEVWGIPIIDVTFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGSLTQOI 1109
QY 1201 VAALSTAKYKNWPESVYFFKRGSGSIQOYIYKQEPVQKCPGRRPALNYPVYGETTQVRRR 1260
DB 1110 VAALSTAKYKNWPESVYFFKRGSGSIQOYIYKQEPVQKCPGRRPALNYPVYGETTQVRRR 1169
QY 1261 PERAIGSQTHIRIQXSPARLAYQDKGVLHNEVKVSIILWRGLENNVTSALSLENIRKPD 1320
DB 1170 PERAIGSQTHIRIQXSPARLAYQDKGVLHNEVKVSIILWRGLENNVTSALSLENIRKPD 1229
QY 1321 GYDYAFSKOQYINIDVPSRTARAITTRSGQTLSKVMYNCP 1361
DB 1230 GYDYAFSKOQYINIDVPSRTARAITTRSGQTLSKVMYNCP 1270
RESULT 10
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX AC AAM24322;
XX AC
DT 12-OCT-2001 (first entry)
XX DE
DE Human EST encoded protein SEQ ID NO: 1847.
XX KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
XX KW Gene therapy; nutrition.
OS Homo sapiens.
XX OS
XX WO200154477-A2.
XX PN
XX 02-AUG-2001.
XX 25-JAN-2001; 2001WO-US002687.
XX 25-JAN-2001; 2001WO-US002687.
PR 25-JAN-2001; 2001WO-US002687.
PR 17-JUL-2000; 2000US-00491404.
PR 03-AUG-2000; 2000US-00617746.
PR 15-SEP-2000; 2000US-00631451.
XX 15-SEP-2000; 2000US-00663870.
XX (HYSE-) HYSBQ INC.
XX PA
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX WPI; 2001-476164/51.
DR N-PSDB; AAH98981.
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX Claim 20; Page 1198-1201; 1275pp; English.
XX The present invention provides the protein and coding sequences of novel
CC

CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention
XX
SQ Sequence 1299 AA;

Query Match 91.9%; Score 6732.7; DB 4; Length 1299;
Best Local Similarity 96.7%; Pred. No. 5.9e-162;
Matches 1252; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

Qy 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCQYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCQYNCQHYMECCPDF 60
Qy 61 KRVTABLSCKGRCFESFERGECDCDAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
Db 61 KRVTABLSCKGRCFESFERGECDCDAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
Qy 121 PPGASQTIKSTTKRSKPPNKKTKKVIIESEITE----- 156
Db 121 PPGASQTIKSTTKRSKPPNKKTKKVIIESEITEHSVSENQESSSSSSSSSTI 180
Qy 157 -----VKDNKNRTKKKPTPKPPVVDVAGSLDNGDFKVTTPDTST 197
Db 181 KIKSKNSAANRELQKLKVDKNKNRTKKKPTPKPPVVDVAGSLDNGDFKVTTPDTST 240
Qy 198 TOHNKVSPTKITAKPNRPSPNSDTSKETSIVNKETTVETKTTTTNKQTSIDG 257
Db 241 TOHNKVSPTKITAKPNRPSPNSDTSKETSIVNKETTVETKTTTTNKQTSIDG 300
Qy 258 KEKTSASQTSIEKTSADLAPTSKVLAKPTKAEITTKGPALTTPKKEPTTPPKSPAS 317
Db 301 KEKTSASQTSIEKTSADLAPTSKVLAKPTKAEITTKGPALTTPKKEPTTPPKSPAS 360
Qy 318 TTPKEPTTTIKSAPTTKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 377
Db 361 TTPKEPTTTIKSAPTTKPEAPTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 420
Qy 378 APPTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 437
Db 421 APPTTKSAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 480
Qy 438 EPAPTAPKKPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 497
Db 481 EPAPTAPKKPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 540
Qy 498 TTGSAPTTKPSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 557
Db 541 TTGSAPTTKPSPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 600
Qy 558 APAPKPEAPTTKPEAPTTKPKLPTTPEKLAFTTPEELAPTTPEELAPTTPEPTT 617
Db 601 APAPKPEAPTTKPEAPTTKPKLPTTPEKLAFTTPEELAPTTPEELAPTTPEPTT 660
Qy 618 PEEAPPTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 677
Db 661 PEEAPPTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPT 720
Qy 678 APPTPKKAPKELAPTTKPEPTSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTKPG 737
Db 721 APPTPKKAPKELAPTTKPEPTSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTKPG 780
Qy 738 TAPPTLKPEAPTTKPKAPKELAPTTTKGPTSTSDKAPTTPKETATTTKPEAPTTPK 797
Db 781 TAPPTLKPEAPTTKPKAPKELAPTTTKGPTSTSDKAPTTPKETATTTKPEAPTTPK 840
Qy 798 KPAPPTTPTPTTSEVSTPTTTKGTTHKSPDSTBELSAEPTPKALENSPKFPGVPT 857
Db 841 KPAPPTTPTPTTSEVSTPTTTKGTTHKSPDSTBELSAEPTPKALENSPKFPGVPT 900

Qy 858 TKTPAATKPEMTTAKDKTTTERDLRTTPTTAAPKMTKETATTTTEKTESKITATTQV 917
Db 901 TKTPAATKPEMTTAKDKTTTERDLRTTPTTAAPKMTKETATTTTEKTESKITATTQV 960
Qy 918 TSITTTQDTTTPFKITTLTKTTTLAPKVTTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 977
Db 961 TSITTTQDTTTPFKITTLTKTTTLAPKVTTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020
Qy 978 PQKETKAPKPTSTKPKTMRVRKPKTTPPRKMTSTMBELNPTSRISAEAMLQTTTRPN 1037
Db 1021 PQKETKAPKPTSTKPKTMRVRKPKTTPPRKMTSTMBELNPTSRISAEAMLQTTTRPN 1080
Qy 1038 QTPNSKLIVNPKSEDAGGAEGETHMLLRPHVFMPEVTPDMDVLRPVNPGIINPMLS 1097
Db 1081 QTPNSKLIVNPKSEDAGGAEGETHMLLRPHVFMPEVTPDMDVLRPVNPGIINPMLS 1140
Qy 1098 DETNINCKGKPVVDGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSIDTVFT 1157
Db 1141 DETNINCKGKPVVDGLTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSIDTVFT 1200
Qy 1158 RCNCEGKTPFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKWPESVY 1217
Db 1201 RCNCEGKTPFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKWPESVY 1260
Qy 1218 FFKRGGSIQYIYKQEPVQKCPGRPALNYPVYGE 1252
Db 1261 FFKRGGSIQYIYKQEPVQKCPGRPALNYPVYGE 1295

RESULT 11
RAB29778
ID AAB29778 standard; protein; 902 AA.
XX
AC AAB29778;
XX
DT 28-FEB-2001 (first entry)
XX
DE Human MSP-derived tribonectin.
XX
KW Human tribonectin; MSP; megakaryocyte stimulating factor;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX
OS Homo sapiens.
XX
PN WO200064930-A2.
XX
PD 02-NOV-2000.
XX
PF 24-APR-2000; 2000WO-US010953.
XX
PR 23-APR-1999; 99US-00298970.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Jay GD;
XX
DR WPI; 2001-024673/03.
XX
PT Novel tribonectin polypeptide useful as lubricant for treating
XX osteoarthritis, comprises O-linked lubricating moiety.
PS Disclosure; Fig 1; 47pp; English.
XX
CC The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSP (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPTT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSP-derived tribonectin; a biocompatible composition comprising a human

CC tribonectin for inhibiting tissue adhesion formation; and a method of
 CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
 CC measuring the amount of MSF or its fragment in a biological sample of a
 CC mammal, wherein an increased amount of MSF compared to a control
 CC indicates the presence of or predisposition to developing osteoarthritis.
 CC The tribonectin and DNA encoding it are useful in the treatment of
 CC osteoarthritis, where they may be used for lubricating mammalian joints,
 CC such as articulating joints of humans, dogs or horses. The tribonectin,
 CC when formulated as a membrane, foam, gel or fibre, is useful for
 CC inhibiting adhesion between two surfaces such as the injured tissues of a
 CC mammal, where the injury is caused by a surgical insertion or trauma, or
 CC an artificial device e.g., an orthopaedic implant. In particular, one of
 CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
 CC used in gene therapy. The present sequence represents a substantial
 CC portion of a human MSF-derived tribonectin
 XX
 XX Sequence 902 AA;

Query Match 48.5%; Score 3552; DB 4; Length 902;
 Best Local Similarity 74.4%; Pred. No. 1e-81;
 Matches 795; Conservative 27; Mismatches 76; Indels 170; Gaps 66;

QY 157 VKONKQRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPIN 216
 Db 1 VKONKQRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPIN 60
 QY 217 PRPSLPNSDTSKETSJTVNKETTVETKETTNNKQSTDGKEKTTSAKETQSTIEKTSK 276
 Db 61 PRPSLPNSDTSKETSJTVNKETTVETKETTNNKQSTDGKEKTTSAKETQSTIEKTSK 120
 QY 277 DLAPTSKVLAKPTPKAETTTKGPALTTTPKPEPTTPPKPEASTTPKPEPTTPTTKSAPTPK 336
 Db 121 DLAPTSKVLAKPTPKAETTTKGPALTTTPKPEASTTPKPEPTTPTTKSAPTPK 172
 QY 337 EPAPTTTKSAPTPPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTT 396
 Db 173 EPAPTTTKSAPTPPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 228
 QY 397 PKPEAPTTKPEAPTTTKPEPTTPPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTKPEA 456
 Db 229 -KEAPTTT-KEAPTTT-KEAPTTT-KEAPTTTKEAPTTT-KEAPTTT-KEAPTTT-KEA 281
 QY 457 PTTKPEAPTTTKPEPTTPPKPEAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEPTTKKE 516
 Db 282 PTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTTKE 335
 QY 517 PAPTTPKEAPPTPKKPAPTTPKPEAPTTTPKPEAPTTTKKPAPTAPKPEAPTTTKETAPT 576
 Db 336 PAPTTP-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTT 388
 QY 577 TPKKLTPPTPEKLAPTTPEKPAPTTPELAPTTPEEPTTPTEEPAPTTTPKAAAPNTPKE 636
 Db 389 T-KEAPAPTTKEP-APTTPKE-PAPTTP-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEA 440
 QY 637 PAPTTPKEAPPTTKPEAPTTTKETAPTTPKGTAPTTLKEAPPTTPKPKAPKELAPTTTK 696
 Db 441 PAPTTP-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTT 488
 QY 697 EPTSTSDKPAPTTPKGTAPTTPKPEAPTTTPKGPAPTTLKEAPPTTPKPKAP 756
 Db 489 EPAPTTPKE-PAPTTP-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTTT-KEAPAPTT 541
 QY 757 --KELAPTTTKGTSTSDKPAPTTPKETAPTTPKPEAPTTTPKPKAPTTPEITPPPTTSEV 814
 Db 542 TTKEPAP-TTKPEAPTTTKE-PAPTTP-KEAPAPTTT-KEAPAPTTT-KEAPAPTTTKE-PAPTTPKEP 595
 QY 815 STPTTKETPTTIHKSDPESTPELSAEPKALENSP--KEPGVPTTKTPAATKPEMTTIA 872
 Db 596 A--PTTKEPAPTTTKEPAPTTT---KEPAPTTTKEPAPTTTKEP-APTTPKEAPPTT----- 640
 QY 873 KOKITERDURTTPETTTAAAPKMTKERATTTTEKTESKITATTTQVSTTTTQDTPPKITT 932
 Db 641 -----TKEAPAPTTT-----KEPAPTTTKEPAPTTTKEPAPTT----- 667

QY 933 LKTTTLAPKVTTTTKKTTTITTEIMNKPEETAKPKDRATNASKATTPKPKPTKAPKKPTSTK 992
 Db 668 -----TTPKEAPTTTKEP-APT-TK 684
 QY 993 KPKTMPRVKPKTTTPTPRKMTSTMPPELNPSTSRIAEAMLQTTTRNQTNPNSKLVENPKSE 1052
 Db 685 EP-----APTTPTRKMTSTMPPELNPSTSRIAEAML-TTTRNQTNPNSKLVENPKSE 735
 QY 1053 DAGGAEGETPHMLLRPHVFMPEVTPDMDYLPFRVFNQGIINPMLSDETNICNGKXPVDGLT 1112
 Db 736 DAGGAEGETPHMLLRPHVFMPEVTPDMDYLPFRVFNQGIINPMLSDETNICNGKXPVDGLT 795
 QY 1113 TLRNGTLVAFRGHYFWMLSFPSPSPARRITVEWGIPIPSIDTTPTRCNCSGKTFFFKDSQ 1172
 Db 796 TLRNGTLVAFRGHYFWMLSFPSPSPARRITVEWGIPIPSIDTTPTRCNCSGKTFFFKDSQ 855
 QY 1173 YWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFK 1220
 Db 856 YWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTA-YKNWPESVYFFK 902

RESULT 12
 ABUS3253

ID ABUS3253 standard; protein; 551 AA.

XX ABUS3253;

XX DT 14-APR-2003 (first entry)

XX DE Human testes-derived DKFZpthes3_4019 homologue #2.

XX KW Human; gene therapy; vaccine; disease treatment; detection.

XX OS Homo sapiens.

XX PN WO200112659-A2.

XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000WO-IB001496.

XX PR 18-AUG-1999; 99US-0149499P.

XX PR 28-SEP-1999; 99US-0156503P.

XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX PI Wiemann S;

XX XX WPI; 2001-327840/34.

XX Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.

XX Example III; Page 892-893; 1095pp; English.

CC This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

XX Sequence 551 AA;

Query Match 40.0%; Score 2929; DB 4; Length 551;
 Best Local Similarity 100.0%; Pred. No. 2.9e-66;
 Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 554 TKKPAPAKPAPPTPKETAPTPKKTTPTEKLAAPTPKEPAPTPPELAPPTPEEP 613
XX |||||
Db 1 TKKPAPAPKAPAPPTPKETAPTPKKTTPTEKLAAPTPKEPAPTPPELAPPTPEEP 60
QY 614 TPTTPEPAPPTPKAAAPNTPKAPAPPTPKAPAPPTPKAPAPPTPKETAPTPKGTAPTT 673
XX |||||
Db 61 TPTTPEPAPPTPKAAAPNTPKAPAPPTPKAPAPPTPKAPAPPTPKETAPTPKGTAPTT 120
QY 674 LKBPAPTPKKAPKELAPPTTKETPTSTSDKAPPTPKGTAPTPKEPAPTPKEPAPT 733
XX |||||
Db 121 LKBPAPTPKKAPKELAPPTTKETPTSTSDKAPPTPKGTAPTPKEPAPTPKEPAPT 180
QY 734 TPKGAPTTLKBPAPTPPKAPKELAPPTTKGTSTSDKAPPTPKETAPTPKEPAP 793
XX |||||
Db 181 TPKGAPTTLKBPAPTPPKAPKELAPPTTKGTSTSDKAPPTPKETAPTPKEPAP 240
QY 794 TTPKKAPPTTPPPPTTSVSTPTTKETPTTIHKSPDESTPELSAEPPTKALENSPKP 853
XX |||||
Db 241 TTPKKAPPTTPPPPTTSVSTPTTKETPTTIHKSPDESTPELSAEPPTKALENSPKP 300
QY 854 GVTPTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETTATTEKTTESKITAT 913
XX |||||
Db 301 GVTPTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETTATTEKTTESKITAT 360
QY 914 TTQVSTTTQDTPPEKITTLTKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKA 973
XX |||||
Db 361 TTQVSTTTQDTPPEKITTLTKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKA 420
QY 974 TTPKPKPTKAPKKPTSTKKPKIMPRVRKPTTTPPKMTSTMPELNPTSRIAEAMLQTT 1033
XX |||||
Db 421 TTPKPKPTKAPKKPTSTKKPKIMPRVRKPTTTPPKMTSTMPELNPTSRIAEAMLQTT 480
QY 1034 TRNQTPNSKLVNPKSEBAGGEGTPEMLLRPHVFMPEVTPDMDYLRVFNQGIIN 1093
XX |||||
Db 481 TRNQTPNSKLVNPKSEBAGGEGTPEMLLRPHVFMPEVTPDMDYLRVFNQGIIN 540
QY 1094 PMLSDETNICN 1104
XX |||||
Db 541 PMLSDETNICN 551
```

RESULT 13

ABU53252

ID ABU53252 standard; protein; 546 AA.

XX

AC ABU53252;

XX

DT 14-APR-2003 (first entry)

XX

XX Human testes-derived DKFZphtes3_4o19 homologue #1.

DE

XX Human; gene therapy; vaccine; disease treatment; detection.

KW

XX Homo sapiens.

OS

XX WO200112659-A2.

PN

XX 22-FEB-2001.

FD

XX 18-AUG-2000; 2000WO-IB001496.

XX

XX 18-AUG-1999; 99US-0149499P.

PR

XX 28-SEP-1999; 99US-0156503P.

XX

XX (GEHU-) GERMAN HUMAN GENOME PROJECT.

FA

XX Wiemann S;

PI

XX WPI; 2001-327840/34.

XX

XX Nucleic acids having the sequences of clones isolated from libraries of

PT different human tissues, useful in recombinant DNA methodologies.

PT

XX

PS Example III; Page 892; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

XX Sequence 546 AA;

SQ Query Match 39.9%; Score 2920; DB 4; Length 546;

Best Local Similarity 100.0%; Pred. No. 4.8e-66;

Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 KKBPATTPKEPAPPTPKAPAPPTTKPESPPTPKAPAPPTTKSAPTTTKAPTTKSGAPT 504

Db 1 KKBPATTPKEPAPPTPKAPAPPTTKPESPPTPKAPAPPTTKSAPTTTKAPTTKSGAPT 60

QY 505 TPKEPSPTTKPEPAPPTPKAPAPPTTPKKAPAPPTPKAPAPPTTKKAPAPTPAPKE 564

Db 61 TPKEPSPTTKPEPAPPTPKAPAPPTTPKKAPAPPTPKAPAPPTTKKAPAPTPAPKE 120

QY 565 PAPTTPKEPAPTPKLTPTTPEKLAAPTTPKEPAPTPPELAPTTPEEPTTPEEPAPT 624

Db 121 PAPTTPKEPAPTPKLTPTTPEKLAAPTTPKEPAPTPPELAPTTPEEPTTPEEPAPT 180

QY 625 TPXAAAPNTPKAPAPPTPKAPAPPTPKAPAPPTPKETAPTPKGTAPTTLKEPAPTPKK 684

Db 181 TPXAAAPNTPKAPAPPTPKAPAPPTPKAPAPPTPKETAPTPKGTAPTTLKEPAPTPKK 240

QY 685 PAPKELAPTTTKETPTSTSDKAPPTPKGTAPTPKEPAPTPPKAPAPTPKGTAPTTLK 744

Db 241 PAPKELAPTTTKETPTSTSDKAPPTPKGTAPTPKEPAPTPPKAPAPTPKGTAPTTLK 300

QY 745 EPAPTTPKAPKELAPTTTKGPTSTSDKAPAPTPPKETAPTPKEPAPTPPKAPAPTPP 804

Db 301 EPAPTTPKAPKELAPTTTKGPTSTSDKAPAPTPPKETAPTPKEPAPTPPKAPAPTPP 360

QY 805 ETPEPTTSEVSTPTTKETPTTIHKSPDESTPELSAEPPTKALENSPKPEGVPTTKTAAAT 864

Db 361 ETPEPTTSEVSTPTTKETPTTIHKSPDESTPELSAEPPTKALENSPKPEGVPTTKTAAAT 420

QY 865 KPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETTATTEKTTESKITATTTQVSTTTQD 924

Db 421 KPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETTATTEKTTESKITATTTQVSTTTQD 480

QY 925 TTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 984

Db 481 TTPPKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 540

QY 985 PKKPTS 990

Db 541 PKKPTS 546

RESULT 14

AAO18834

ID AAO18834 standard; protein; 538 AA.

XX

AC AAO18834;

XX

XX 29-OCT-2002 (first entry)

DT

XX 3' cartilage superficial zone protein coding sequence encoded protein.

DE

XX SZP; superficial zone protein; cartilage; lubrication; human;

KW degenerative joint condition; arthritis; osteoporosis; trauma; CACP;

KW chondroitin sulphate substitution consensus; antiarthritic;

KW antirheumatic; osteopathic; antigout; antiinflammatory; dermatological;
 KW immunosuppressive.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH Misc-difference 114
 FT /note= "encoded by ACTACT"
 XX
 XX WO200262847-A2.
 XX
 XX 15-AUG-2002.
 XX
 XX 31-DEC-2001; 2001WO-US050379.
 XX
 XX 29-DEC-2000; 2000US-0258920P.
 XX
 XX (GLAXO) GLAXO GROUP LTD.
 XX (RUSH-) RUSH PRESBYTERIAN ST LUKE MEDICAL CENT.
 XX Hutchins JT, Kuettnner KE, Schmid TM, Schumacher BL, Su J;
 PI Dixon EP;
 XX
 XX WPI; 2002-636585/68.
 DR N-PSDB; AAL49079.
 XX
 XX New purified superficial zone protein (SZP) polypeptides, useful for
 PT treating degenerative joint conditions, e.g. osteoarthritis, rheumatoid
 PT arthritis, gout, spondylarthritis, synovitis, tendonitis, lupus, or
 PT osteoporosis.
 XX
 XX Claim 59; Page 86-87; 89pp; English.
 XX
 XX The present invention provides the protein and coding sequences of human
 CC superficial zone protein (SZP). The protein is involved in the
 CC lubrication of joints, and the sequences can be used in the treatment of
 CC degenerative joint conditions or to delay symptoms of a degenerative
 CC joint condition, e.g. osteoarthritis, rheumatoid arthritis, gout,
 CC psoriatic arthritis, reactive arthritis, viral or post viral arthritis,
 CC spondylarthritis, juvenile arthritis, synovitis, tendonitis, systemic
 CC lupus erythematosus, CACP, osteoporosis or trauma. The present sequence
 CC is the protein encoded by the human 3' cartilage SZP cDNA
 XX
 XX Sequence 538 AA;
 Query Match 38.9%; Score 2850.9; DB 5; Length 538;
 Best Local Similarity 99.8%; Pred. No. 2.6e-64;
 Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 823 PTTIHKSPESTPELSAEPFKALENSPKPEPGVPTTKPAATKPEMTTAKDKTTERDLR 882
 Db 1 PTTIHKSPESTPELSAEPFKALENSPKPEPGVPTTKPAATKPEMTTAKDKTTERDLR 60
 QY 883 TTPETTAAPKMKETATTTTEKTESKITATTTQVSTTTQDTPPKITLKTTLAPKV 942
 Db 61 TTPETTTAAPKMKETATTTTEKTESKITATTTQVSTTTQDTPPKITLKTTLAPKV 119
 QY 943 TTTTKKITTITEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKPTSTKKPKTMRVRK 1002
 Db 120 TTTTKKITTITEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKPTSTKKPKTMRVRK 179
 QY 1003 PKTTPPRKMTSTWPELNPSTRIAEAMLOTTTPRNPQNSKLEVPNPKSDAGAGETP 1062
 Db 180 PKTTPPRKMTSTWPELNPSTRIAEAMLOTTTPRNPQNSKLEVPNPKSDAGAGETP 239
 QY 1063 HMLLRPHVFPEVTPDMDYLPVRPNQGIINPMLSDETNICNGKPDGLTTLRNGTLVAF 1122
 Db 240 HMLLRPHVFPEVTPDMDYLPVRPNQGIINPMLSDETNICNGKPDGLTTLRNGTLVAF 299
 QY 1123 RGHYFWMLSFSPSPARRITEVWGIPTSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKD 1182
 Db 300 RGHYFWMLSFSPSPARRITEVWGIPTSPIDTVFTRCNCEGKTFFFKDSQYWRFTNDIKD 359

QY 1183 AGYPKPIFKGFGGLTGQIIVAALSTAKYKNWPESVYFFKRGSIQQYIYKQEPVQKCPGRR 1242
 Db |||||
 QY 360 AGYPKPIFKGFGGLTGQIIVAALSTAKYKNWPESVYFFKRGSIQQYIYKQEPVQKCPGRR 419
 Db |||||
 QY 1243 PALNYPVYGEMTQVRRRRFERRAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRG 1302
 Db |||||
 QY 420 PALNYPVYGEMTQVRRRRFERRAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRG 479
 Db |||||
 QY 1303 LPNVVTSALSLPNIRKPDGYDYAFSKDQYINIDVFSRTARAITTRSGQTLKQWVNC 1361
 Db |||||
 QY 480 LPNVVTSALSLPNIRKPDGYDYAFSKDQYINIDVFSRTARAITTRSGQTLKQWVNC 538
 Db |||||
 RESULT 15
 ABUS3254
 ID ABUS3254 standard; protein; 513 AA.
 XX
 XX AC ABUS3254;
 XX
 XX 14-APR-2003 (first entry)
 XX
 XX Human testes-derived DKFZphtes3_4019 homologue #3.
 DE Human; gene therapy; vaccine; disease treatment; detection.
 XX Homo sapiens.
 KW
 KW Human
 OS
 XX
 XX WO200112659-A2.
 PN
 XX 22-FEB-2001.
 PD
 XX
 XX 18-AUG-2000; 2000WO-IB001496.
 PF
 XX
 XX 18-AUG-1999; 99US-0149499P.
 PR
 XX 28-SEP-1999; 99US-0156503P.
 PR
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
 PA
 XX Wiemann S;
 PI
 XX WPI; 2001-327840/34.
 DR
 XX
 XX Nucleic acids having the sequences of clones isolated from libraries of
 PT different human tissues, useful in recombinant DNA methodologies.
 PT
 XX
 XX Example III; Page 893; 1095pp; English.
 PS
 XX
 CC This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention
 XX
 XX Sequence 513 AA;
 Query Match 37.6%; Score 2757; DB 4; Length 513;
 Best Local Similarity 100.0%; Pred. No. 5.8e-62;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 164 RTKKXPTPKPPVVDEAGSLDNGDFKVTTPDTSSTQHNKVSTSPKITTAKPINPRESLPP 223
 Db 1 RTKKXPTPKPPVVDEAGSLDNGDFKVTTPDTSSTQHNKVSTSPKITTAKPINPRESLPP 60
 QY 224 NSDTSKTSLSLVNKKETTVETKETTINKQSTDKCKTTSKAKTQSIKTSKDLAPTSK 283
 Db 61 NSDTSKTSLSLVNKKETTVETKETTINKQSTDKCKTTSKAKTQSIKTSKDLAPTSK 120
 QY 284 VLAKPTPKAETTTKGPALTPPKPTPTTPKEPASTTPKPTTTIKSAPTTPKEPAPTTT 343

```
Db      121  VLAKPTPKAETTTKGPALTTPKERTTPKKEPASTTPKERTPTTIKSAPTTPKEPAPTTT 180
QY      344  KSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTPKKDAPT 403
Db      181  KSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTPKKDAPT 240
QY      404  TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTAPKKDAPTTPKEPAPTTTPKEP 463
Db      241  TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTAPKKDAPTTPKEPAPTTTPKEP 300
QY      464  APTTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPK 523
Db      301  APTTTKEPSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPK 360
QY      524  EPAPTTPKKDAPTTPKEPAPTTTPKEPAPTTTKKDPAPTAPKEPAPTTTPKEPAPTTTPKCLTP 583
Db      361  EPAPTTPKKDAPTTPKEPAPTTTPKEPAPTTTKKDPAPTAPKEPAPTTTPKEPAPTTTPKCLTP 420
QY      584  TTPEKLAPTTPKEPAPTTPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTPKAAAAPNTKGPAPTTPK 643
Db      421  TTPEKLAPTTPKEPAPTTPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTPKAAAAPNTKGPAPTTPK 480
QY      644  EPAPTTPKEPAPTTPKETAPTTTPKGTAPTTIKE 676
Db      481  EPAPTTPKEPAPTTPKETAPTTTPKGTAPTTIKE 513
```

Search completed: October 13, 2004, 11:37:22
Job time : 116.728 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 23.6582 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-G

Perfect score: 7323

Sequence: 1 MAWKLPYVLLLSVFIQ.....ARAITRSQGTLKSVYNCP 1361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 7323 | 100.0 | 1361 | 4 | US-07-757-022B-40 |
| 2 | 7308.7 | 99.8 | 1404 | 4 | US-07-757-022B-2 |
| 3 | 7308.7 | 99.8 | 1404 | 4 | US-07-757-022B-62 |
| 4 | 7308.7 | 99.8 | 1404 | 4 | US-09-298-970A-1 |
| 5 | 7293.7 | 99.6 | 1404 | 4 | US-10-164-595-78 |
| 6 | 7084.3 | 96.7 | 1354 | 4 | US-07-757-022B-48 |
| 7 | 7058.9 | 96.4 | 1320 | 4 | US-07-757-022B-46 |
| 8 | 7058.9 | 96.4 | 1320 | 4 | US-07-757-022B-60 |
| 9 | 7049 | 96.3 | 1311 | 4 | US-07-757-022B-42 |
| 10 | 7044.6 | 96.2 | 1363 | 4 | US-07-757-022B-52 |
| 11 | 7043.9 | 96.2 | 1320 | 4 | US-10-164-595-58 |
| 12 | 6836.3 | 93.4 | 1314 | 4 | US-07-757-022B-50 |
| 13 | 6820.2 | 93.1 | 1313 | 4 | US-07-757-022B-142 |
| 14 | 6784.9 | 92.7 | 1270 | 4 | US-07-757-022B-44 |
| 15 | 5867.7 | 80.1 | 1140 | 4 | US-07-757-022B-104 |
| 16 | 5485.7 | 74.9 | 1038 | 4 | US-07-757-022B-74 |
| 17 | 5472 | 74.7 | 1022 | 4 | US-07-757-022B-84 |
| 18 | 5379.2 | 73.5 | 1049 | 4 | US-07-757-022B-58 |
| 19 | 5011 | 68.4 | 941 | 4 | US-07-757-022B-14 |
| 20 | 2246.2 | 30.7 | 463 | 4 | US-07-757-022B-54 |
| 21 | 1891.2 | 27.2 | 423 | 4 | US-07-757-022B-66 |
| 22 | 1882.1 | 27.1 | 422 | 4 | US-07-757-022B-68 |
| 23 | 1723.1 | 23.5 | 372 | 4 | US-07-757-022B-64 |
| 24 | 1413.6 | 19.3 | 5179 | 4 | US-09-538-032-1258 |
| 25 | 1295.5 | 17.7 | 296 | 4 | US-07-757-022B-70 |
| 26 | 1214.6 | 16.6 | 237 | 4 | US-07-757-022B-72 |
| 27 | 1189.7 | 16.2 | 8991 | 4 | US-08-714-741-32 |

Sequence 116, App
Sequence 136, App
Sequence 96, Appl
Sequence 30, Appl
Sequence 32, Appl
Sequence 94, Appl
Sequence 132, App
Sequence 92, Appl
Sequence 90, Appl
Sequence 88, Appl
Sequence 1, Appl
Sequence 106, App
Sequence 118, App
Sequence 102, App
Sequence 114, App
Sequence 2, Appl
Sequence 98, Appl
Sequence 22, Appl

28 1143 15.6 207 4 US-07-757-022B-116
29 1143 15.6 207 4 US-07-757-022B-136
30 970.7 13.3 220 4 US-07-757-022B-96
31 956 13.1 231 4 US-07-757-022B-30
32 918.7 12.5 188 4 US-07-757-022B-32
33 908.7 12.4 209 4 US-07-757-022B-94
34 903.7 12.3 208 4 US-07-757-022B-132
35 892.1 12.2 204 4 US-07-757-022B-92
36 885.7 12.1 192 4 US-07-757-022B-90
37 875.9 12.0 172 4 US-07-757-022B-88
38 872.6 11.9 3118 3 US-09-579-181-1
39 871 11.9 156 4 US-07-757-022B-106
40 871 11.9 156 4 US-07-757-022B-118
41 869 11.9 157 4 US-07-757-022B-102
42 869 11.8 2972 3 US-09-579-181-2
43 862.9 11.7 3256 4 US-09-919-172-98
44 860.1 11.7 3256 4 US-09-976-594-22
45 860.1 11.7 3256 4 US-09-976-594-22

US-07-757-022B-40
; Sequence 40, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID

ALIGNMENTS

| | | |
|---|------|---|
| ; TOPOLOGY: linear | | |
| ; MOLECULE TYPE: protein | | |
| US-07-757-022B-40 | | |
| Query Match 100.0%; Score 7323; DB 4; Length 1361; | | |
| Best Local Similarity 100.0%; Pred. No. 2e-199; | | |
| Matches 1361; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | |
| Qy | 1 | MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEYSRDATCNCYINCOHYMECCPDF 60 |
| Db | 1 | MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEYSRDATCNCYINCOHYMECCPDF 60 |
| Qy | 61 | KRVCTAELCKGRFCFESFERGECDDAOCKKYDKCCPDYFSCAEVHNPTSPSSKKAP 120 |
| Db | 61 | KRVCTAELCKGRFCFESFERGECDDAOCKKYDKCCPDYFSCAEVHNPTSPSSKKAP 120 |
| Qy | 121 | PPGASQTIKSTTKRSKPPNKKTKVIESEETEVKDNKNKTKKKTPKPPVDEAG 180 |
| Db | 121 | PPGASQTIKSTTKRSKPPNKKTKVIESEETEVKDNKNKTKKKTPKPPVDEAG 180 |
| Qy | 181 | SGLDNGDFKVTTPDTSITQHNKVSPTSPIITAKINPRPSLPNSDTSKETSIVNKETT 240 |
| Db | 181 | SGLDNGDFKVTTPDTSITQHNKVSPTSPIITAKINPRPSLPNSDTSKETSIVNKETT 240 |
| Qy | 241 | VEKTETTTTNKQTSIDGKEKITSKETSISAKDLAPTSKVLAKPTPKAETTTKGA 300 |
| Db | 241 | VEKTETTTTNKQTSIDGKEKITSKETSISAKDLAPTSKVLAKPTPKAETTTKGA 300 |
| Qy | 301 | LTPPKETPTTPKPAETTKPEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 360 |
| Db | 301 | LTPPKETPTTPKPAETTKPEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 360 |
| Qy | 361 | PAPTTKPEAPTTTKPEAPTTKSAPTTKPEAPTTTKKPAPTTKPEAPTTKPEPTT 420 |
| Db | 361 | PAPTTKPEAPTTTKPEAPTTKSAPTTKPEAPTTTKKPAPTTKPEAPTTKPEPTT 420 |
| Qy | 421 | PKEAPTTKPEAPTTKPEAPTPAKPAPTTKPEAPTTKPEAPTTTKESPTTKPEA 480 |
| Db | 421 | PKEAPTTKPEAPTTKPEAPTPAKPAPTTKPEAPTTKPEAPTTTKESPTTKPEA 480 |
| Qy | 481 | PTTTKSAPTTTKPEAPTTKSAPTTKPEPSPTTKPEAPTTKPEAPTTPKKPAPTTPKE 540 |
| Db | 481 | PTTTKSAPTTTKPEAPTTKSAPTTKPEPSPTTKPEAPTTKPEAPTTPKKPAPTTPKE 540 |
| Qy | 541 | PAPTTKPEAPTTTKKPAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 600 |
| Db | 541 | PAPTTKPEAPTTTKKPAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 600 |
| Qy | 601 | TPBELAPTTPEPTPTTPEAPTTKPAAPNTKPEAPTTKPEAPTTKPEAPTTTPKE 660 |
| Db | 601 | TPBELAPTTPEPTPTTPEAPTTKPAAPNTKPEAPTTKPEAPTTKPEAPTTTPKE 660 |
| Qy | 661 | TAPTTKPGTAPTTLKEAPTTPKKPAKELAPTTTKGPTSTSDKPAPTTPKGAPTPPK 720 |
| Db | 661 | TAPTTKPGTAPTTLKEAPTTPKKPAKELAPTTTKGPTSTSDKPAPTTPKGAPTPPK 720 |
| Qy | 721 | EPAPTTKPEAPTTKPGTAPTTLKEAPTTTKKPAKELAPTTTKGPTSTSDKPAPTTP 780 |
| Db | 721 | EPAPTTKPEAPTTKPGTAPTTLKEAPTTTKKPAKELAPTTTKGPTSTSDKPAPTTP 780 |
| Qy | 781 | KETAPTTKPEAPTTPKKPAPTTPETPTPTTSEVSTPTTKETPTTIHKSPESTPELSAE 840 |
| Db | 781 | KETAPTTKPEAPTTPKKPAPTTPETPTPTTSEVSTPTTKETPTTIHKSPESTPELSAE 840 |
| Qy | 841 | PTPKALENSPKPGVPTTKTTPAAKPEMTTAKOTTERDLRTPETTTAAAPKMTKETAT 900 |
| Db | 841 | PTPKALENSPKPGVPTTKTTPAAKPEMTTAKOTTERDLRTPETTTAAAPKMTKETAT 900 |
| Qy | 901 | TTEKTTESKITATTTQVTSITTTQDTPPKITITLKITTLAPKVTTTKITTTTTEIMNKP 960 |
| Db | 901 | TTEKTTESKITATTTQVTSITTTQDTPPKITITLKITTLAPKVTTTKITTTTTEIMNKP 960 |
| Qy | 961 | TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTTTPRKMTSTMPBLN 1020 |
| Db | 961 | TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTTTPRKMTSTMPBLN 1020 |
| Qy | 1021 | PTSRIAEMLOTTTRPNQTPNSKLVEVNPKSEDAGGAETPHMLLRPHVFMPEVTPDMD 1080 |
| Db | 1021 | PTSRIAEMLOTTTRPNQTPNSKLVEVNPKSEDAGGAETPHMLLRPHVFMPEVTPDMD 1080 |
| Qy | 1081 | YLPRVFNQGIINPMLSDETNINCGRKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPAR 1140 |
| Db | 1081 | YLPRVFNQGIINPMLSDETNINCGRKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPAR 1140 |
| Qy | 1141 | RITEVWGIPIPIDTVFTRCNCCKGKTFKFKDSQYWRFTNDIKDAGYKPIFKFGGLTQCI 1200 |
| Db | 1141 | RITEVWGIPIPIDTVFTRCNCCKGKTFKFKDSQYWRFTNDIKDAGYKPIFKFGGLTQCI 1200 |
| Qy | 1201 | VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQBPVQKCPGRRPALNIPVYGMTQVRRR 1260 |
| Db | 1201 | VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQBPVQKCPGRRPALNIPVYGMTQVRRR 1260 |
| Qy | 1261 | FERAIGPSQHTTIRIQYSPARLAYQDKGVHLHNEVKVSIILWRGLPNVVTSAISLENIRKPD 1320 |
| Db | 1261 | FERAIGPSQHTTIRIQYSPARLAYQDKGVHLHNEVKVSIILWRGLPNVVTSAISLENIRKPD 1320 |
| Qy | 1321 | GYDYAFSKQYQYNNIDVPSRTARAITTRSGOTLSKWYNCP 1361 |
| Db | 1321 | GYDYAFSKQYQYNNIDVPSRTARAITTRSGOTLSKWYNCP 1361 |
| RESULT 2 | | |
| US-07-757-022B-2 | | |
| ; Sequence 2, Application US/07757022B | | |
| ; Patent No. 6433142 | | |
| ; GENERAL INFORMATION: | | |
| ; APPLICANT: Gesner, Thomas G. | | |
| ; APPLICANT: Clark, Stephen C. | | |
| ; APPLICANT: Turner, Katherine | | |
| ; APPLICANT: Hewick, Rodney M. | | |
| ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors | | |
| ; NUMBER OF SEQUENCES: 143 | | |
| ; CORRESPONDENCE ADDRESS: | | |
| ; ADDRESSEE: Genetics Institute, Inc. | | |
| ; STREET: 87 Cambridgepark Drive | | |
| ; CITY: Cambridge | | |
| ; STATE: Massachusetts | | |
| ; COUNTRY: U.S.A. | | |
| ; ZIP: 02140 | | |
| ; COMPUTER READABLE FORM: | | |
| ; MEDIUM TYPE: Floppy disk | | |
| ; COMPUTER: IBM PC compatible | | |
| ; OPERATING SYSTEM: PC-DOS/MS-DOS | | |
| ; SOFTWARE: PatentIn Release #1.0, Version #1.25 | | |
| ; CURRENT APPLICATION DATA: | | |
| ; APPLICATION NUMBER: US/07757.022B | | |
| ; FILING DATE: 19910910 | | |
| ; CLASSIFICATION: 530 | | |
| ; PRIOR APPLICATION DATA: | | |
| ; APPLICATION NUMBER: US 07/643,502 | | |
| ; FILING DATE: 18-JAN-1991 | | |
| ; PRIOR APPLICATION DATA: | | |
| ; APPLICATION NUMBER: US 07/546,114 | | |
| ; FILING DATE: 29-JUN-1990 | | |
| ; PRIOR APPLICATION DATA: | | |
| ; APPLICATION NUMBER: US 07/457,196 | | |
| ; FILING DATE: 29-DEC-1989 | | |
| ; PRIOR APPLICATION DATA: | | |
| ; APPLICATION NUMBER: US 07/390,901 | | |
| ; FILING DATE: 08-AUG-1989 | | |
| ; ATTORNEY/AGENT INFORMATION: | | |
| ; NAME: Cseri, Luan | | |
| ; REGISTRATION NUMBER: 31,822 | | |
| ; REFERENCE/DOCKET NUMBER: GI 5190 | | |
| ; TELECOMMUNICATION INFORMATION: | | |
| ; TELEPHONE: (617)876-1170 | | |

TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-2

Query Match 99.8%; Score 7308.7; DB 4; Length 1404;
Best Local Similarity 96.9%; Pred. No. 5.4e-199;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLILLLSVFVIQVSSODLSSCAGRGEGYSDRATCNCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLILLLSVFVIQVSSODLSSCAGRGEGYSDRATCNCYNCOHYMECCPDF 60

QY 61 KRVCCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYEFCAEYHNPTSPSSKKAP 120
DB 61 KRVCCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYEFCAEYHNPTSPSSKKAP 120

QY 121 PPSCASQTIKSTTKRSPKPNKKTKKVIIEEITE----- 156
DB 121 PPSCASQTIKSTTKRSPKPNKKTKKVIIEEITE----- 156

QY 157 -----VKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTPTDST 180
DB 157 -----VKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTPTDST 180

QY 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTPTDST 197
DB 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPPVVDGAGSLDNGDFKVTPTDST 197

QY 198 TOHNKVTSPKTIITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQTSIDG 257
DB 198 TOHNKVTSPKTIITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQTSIDG 257

QY 241 TOHNKVTSPKTIITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQTSIDG 300
DB 241 TOHNKVTSPKTIITAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQTSIDG 300

QY 258 KEKTTSAKETQSIIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEPAS 317
DB 258 KEKTTSAKETQSIIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEPAS 317

QY 301 KEKTTSAKETQSIIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEPAS 360
DB 301 KEKTTSAKETQSIIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPKPEPAS 360

QY 318 TTPKEPTTPKPEPASPTTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 377
DB 318 TTPKEPTTPKPEPASPTTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 377

QY 361 TTPKEPTTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 420
DB 361 TTPKEPTTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 420

QY 378 APPTTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 437
DB 378 APPTTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 437

QY 421 APPTTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 480
DB 421 APPTTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 480

QY 438 EPAPTAPKKAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 497
DB 438 EPAPTAPKKAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 497

QY 481 EPAPTAPKKAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 540
DB 481 EPAPTAPKKAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 540

QY 498 TTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 557
DB 498 TTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 557

QY 541 TTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 600
DB 541 TTKSAPTTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 600

QY 558 APAPTAPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 617
DB 558 APAPTAPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 617

QY 601 APAPTAPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 660
DB 601 APAPTAPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 660

QY 618 PEEPAPTTPKAAAPNTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 677
DB 618 PEEPAPTTPKAAAPNTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 677

QY 661 PEEPAPTTPKAAAPNTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 720
DB 661 PEEPAPTTPKAAAPNTKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 720

QY 678 APPTTKKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 737
DB 678 APPTTKKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 737

QY 721 APPTTKKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 780
DB 721 APPTTKKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 780

QY 738 TAPPTTKPEPATTPKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 797
DB 738 TAPPTTKPEPATTPKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 797

QY 781 TAPPTTKPEPATTPKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 840
DB 781 TAPPTTKPEPATTPKPAKPELATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPKPEPATTPK 840

QY 798 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDPESTPELSAETTPKALNSKPEKCVPT 857
DB 798 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDPESTPELSAETTPKALNSKPEKCVPT 857

QY 841 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDPESTPELSAETTPKALNSKPEKCVPT 900
DB 841 KPAPTTPETPPPTTSEVSTTTTKEPTTIHKSDPESTPELSAETTPKALNSKPEKCVPT 900

QY 858 TKTPAATKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETATTTTEKTESKITATTTQV 917
DB 901 TKTPAATKPEMTTAKOKTTERDLRTTPEITTAAPKMTKETATTTTEKTESKITATTTQV 960

QY 918 TSTTTQDTPRFTKLTTLTKTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKAITPK 977
DB 961 TSTTTQDTPRFTKLTTLTKTLAPKVTTKTITTEIMNKPEETAKPKDRATNSKAITPK 1020

QY 978 PQKTKAPKPTSTKPKTMTPRVRKPKTTPPRKMTSTMPELNPTSRIAEAMLOTTTPRN 1037
DB 1021 PQKTKAPKPTSTKPKTMTPRVRKPKTTPPRKMTSTMPELNPTSRIAEAMLOTTTPRN 1080

QY 1038 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1097
DB 1081 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1140

QY 1098 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPMSPSPSPARRITEVWGPSPIDTFT 1157
DB 1141 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPMSPSPARRITEVWGPSPIDTFT 1200

QY 1158 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIIVAALSTAKYKNWPESVY 1217
DB 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIIVAALSTAKYKNWPESVY 1260

QY 1218 FFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRPERRAIGPSQTHTRIQY 1277
DB 1261 FFKGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRPERRAIGPSQTHTRIQY 1320

QY 1278 SPARLAYQDQKGLVHNEVKVSVILWRGLNPNVTSLSLNRKPGDGYVAFSKDOYINIDV 1337
DB 1321 SPARLAYQDQKGLVHNEVKVSVILWRGLNPNVTSLSLNRKPGDGYVAFSKDOYINIDV 1380

QY 1338 PSRTARAITTRSGQTLSSKVMWYNCP 1361
DB 1381 PSRTARAITTRSGQTLSSKVMWYNCP 1404

RESULT 3
US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-62

Query Match          99.8%; Score 7308.7; DB 4; Length 1404;
Best Local Similarity 96.9%; Pred. No. 5.4e-199;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLILLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLILLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF 60
QY 61 KRVTAEISCKGRCFESFERGREGCCDAQCKKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120
DB 61 KRVTAEISCKGRCFESFERGREGCCDAQCKKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120
QY 121 PPGASQTIKSTTRSPKPPNKKTKKVIIESEITE----- 156
DB 121 PPGASQTIKSTTRSPKPPNKKTKKVIIESEITEHSVSENESSSSSSSSSSSTI 180
QY 157 -----VDNKNRRTKKPTPKPVVDVDSGLDNGDFKVTTPDTST 197
DB 181 KIKSSKNSAANRELQKLLKVDKNKNRTKKPTPKPVVDVDSGLDNGDFKVTTPDTST 240
QY 198 TOHNKVSTSPKITAKPINRPPSLPPNSDTSKETSLTVNKEITVETKETTITNKQSTDG 257
DB 241 TOHNKVSTSPKITAKPINRPPSLPPNSDTSKETSLTVNKEITVETKETTITNKQSTDG 300
QY 258 KEKTTSAKETQSIEKTSADKAPTSKVLAKEPTPKAETTTKGPALTTPKKEPTTPPKEPAS 317
DB 301 KEKTTSAKETQSIEKTSADKAPTSKVLAKEPTPKAETTTKGPALTTPKKEPTTPPKEPAS 360
QY 318 TTPKEPTTTIKSAPTTPKKEPAPTTTKSAPTTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 377
DB 361 TTPKEPTTTIKSAPTTPKKEPAPTTTKSAPTTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 420
QY 378 APTTKSAPTTPKKEPAPTTPKKAPTTPKKEPAPTTPKKEPAPTTTKKEPAPTTTKKEPAPTTPK 437
DB 421 APTTKSAPTTPKKEPAPTTPKKAPTTPKKEPAPTTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKK 480
QY 438 EPAPTAPKAPPTPKKEPAPTTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKSAPTTTKKEPAPT 497
DB 481 EPAPTAPKAPPTPKKEPAPTTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKSAPTTTKKEPAPT 540
QY 498 TTKSAPTTPKPPSPPTTKKEPAPTTPKKEPAPTTPKKAPTTPKKEPAPTTTKKEPAPTTTKKPP 557
DB 541 TTKSAPTTPKPPSPPTTKKEPAPTTPKKEPAPTTPKKAPTTPKKEPAPTTTKKEPAPTTTKKPP 600
QY 558 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTT 617
DB 601 APTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTT 660
QY 618 PEEPAPTTPKAAANPTPKKEPAPTTPKKEPAPTTPKETAPTTPKETAPTTPKGTAPTTTLKEP 677
DB 661 PEEPAPTTPKAAANPTPKKEPAPTTPKKEPAPTTPKETAPTTPKETAPTTPKGTAPTTTLKEP 720
QY 678 APTTPKAPKAPKELAPTTTKETSTSDKAPTTTPKGTAPTTPKKEPAPTTPKKEPAPTTTPKG 737
DB 721 APTTPKAPKAPKELAPTTTKETSTSDKAPTTTPKGTAPTTPKKEPAPTTPKKEPAPTTTPKG 780
```

```
QY 738 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKDPAPTTPKETAPTTPKKEPAPTTPK 797
DB 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKDPAPTTPKETAPTTPKKEPAPTTPK 840
QY 798 KPAPTTPETPPPTTSEVSTPTTTTKEPTTIIHKSDESTPELSAETPKALENSPKPEGVPT 857
DB 841 KPAPTTPETPPPTTSEVSTPTTTTKEPTTIIHKSDESTPELSAETPKALENSPKPEGVPT 900
QY 858 TKTPAATKPEMTTAKDKTTERDLRTPPTTTTAAPKMTKETATTTTETKTESKITATTQV 917
DB 901 TKTPAATKPEMTTAKDKTTERDLRTPPTTTTAAPKMTKETATTTTETKTESKITATTQV 960
QY 918 TSTTTQDTPFKITLTKTTLAPKVTTKKTIITTEIMNKPEETAKEPKDRATNSKATTPK 977
DB 961 TSTTTQDTPFKITLTKTTLAPKVTTKKTIITTEIMNKPEETAKEPKDRATNSKATTPK 1020
QY 978 PQKETAAPKXPTSTKPKTTPRVKPKTTPTRKMTSTMPELNPTSIAEAMLQTTTRPN 1037
DB 1021 PQKETAAPKXPTSTKPKTTPRVKPKTTPTRKMTSTMPELNPTSIAEAMLQTTTRPN 1080
QY 1038 QTPNSKLVEVNPKSEDAGGAEETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1097
DB 1081 QTPNSKLVEVNPKSEDAGGAEETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1140
QY 1098 DETNINCKGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPARRITEVWGIPSPIDTVFT 1157
DB 1141 DETNINCKGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPARRITEVWGIPSPIDTVFT 1200
QY 1158 RCNCEGTTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1217
DB 1201 RCNCEGTTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1260
QY 1218 FFKEGGSIQQYIYKQEPVQKCPGRPALNVPYVYEMTQVRRRRRERAIGSOTHTIRIQ 1277
DB 1261 FFKEGGSIQQYIYKQEPVQKCPGRPALNVPYVYEMTQVRRRRRERAIGSOTHTIRIQ 1320
QY 1278 SPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLNPDKPDGYDYAFSKDQVYNIDV 1337
DB 1321 SPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLNPDKPDGYDYAFSKDQVYNIDV 1380
QY 1338 PSRTARAITTRSGQTLSKVMYNCP 1361
DB 1381 PSRTARAITTRSGQTLSKVMYNCP 1404
```

RESULT 4

```
US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-298-970A-1
```

```
Query Match          99.8%; Score 7308.7; DB 4; Length 1404;
Best Local Similarity 96.9%; Pred. No. 5.4e-199;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLILLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLILLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF 60
QY 61 KRVTAEISCKGRCFESFERGREGCCDAQCKKYDKCCPDYBSFCAEVHNPTSPSSKKAP 120
```

| | | | |
|----|------|--|------|
| Db | 61 | KRVCTAELUSCKGRCFESFERGECDCDQCKYDKCCPDYBFCFAEVHNPSPSPSSKCAP | 120 |
| Qy | 121 | PPSGASQIWKSTTKBSPKPNKKTKVIESEIITE | 156 |
| Db | 121 | PPSGASQIWKSTTKBSPKPNKKTKVIESEIITEHSHSVENQBSSESSSSSSSSSTIW | 180 |
| Qy | 157 | -----VKDNKNRNTKKXPTPKPPVVVDVAGSLDNGDPKVTTPDST | 197 |
| Db | 181 | KIKSKNSAANRELQKLKLVKDNKNRNTKKXPTPKPPVVVDVAGSLDNGDPKVTTPDST | 240 |
| Qy | 198 | TOHNKVSTSPKITTAKPINRPSLPDPSGOTSKEFSLTVNKETTVETKTTTTNNQTSVDG | 257 |
| Db | 241 | TOHNKVSTSPKITTAKPINRPSLPDPSGOTSKEFSLTVNKETTVETKTTTTNNQTSVDG | 300 |
| Qy | 258 | KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPAUUTTKPEPTPTTKPEPAS | 317 |
| Db | 301 | KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPAUUTTKPEPTPTTKPEPAS | 360 |
| Qy | 318 | TTPEPTPTTIKSAPTTKPEPAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEP | 377 |
| Db | 361 | TTPEPTPTTIKSAPTTKPEPAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEP | 420 |
| Qy | 378 | APTITKSAPTTKPEPAPTTPKKAPTTPKBAPTTTKPEPTPTTKPEPAPTTTKPEPAPTTTKPEP | 437 |
| Db | 421 | APTITKSAPTTKPEPAPTTPKKAPTTPKBAPTTTKPEPTPTTKPEPAPTTTKPEPAPTTTKPEP | 480 |
| Qy | 438 | EPAPTAKKPAPTTKPEPAPTTKPEPAPTTTKKESPTTKPEPAPTTTKSAPTTTKPEPAPTTKEP | 497 |
| Db | 481 | EPAPTAKKPAPTTKPEPAPTTKPEPAPTTTKKESPTTKPEPAPTTTKSAPTTTKPEPAPTTKEP | 540 |
| Qy | 498 | TTKSAPTTKESPTTKKEPAPTTTKBAPTTTKKKPAPTTTKPEPAPTTTKPEPAPTTTKKAP | 557 |
| Db | 541 | TTKSAPTTKESPTTKKEPAPTTTKBAPTTTKKKPAPTTTKPEPAPTTTKPEPAPTTTKKAP | 600 |
| Qy | 558 | APTAPKPEPAPTTKETAPTTPKKLTPTTPEKLAUPTTPEKAPTTTPEELAPTTTPEEPTPTT | 617 |
| Db | 601 | APTAPKPEPAPTTKETAPTTPKKLTPTTPEKLAUPTTPEKAPTTTPEELAPTTTPEEPTPTT | 660 |
| Qy | 618 | PEEPAPTTPKAAANTPKPEPAPTTKPEPAPTTKPEPAPTTTKETAPTTTKGAPTTTLKEP | 677 |
| Db | 661 | PEEPAPTTPKAAANTPKPEPAPTTKPEPAPTTKPEPAPTTTKETAPTTTKGAPTTTLKEP | 720 |
| Qy | 678 | APTTPKKPAPKELAPTTTKESTSTSDKAPTTTPKGAPTTTKPEPAPTTTKPEPAPTTTKPKG | 737 |
| Db | 721 | APTTPKKPAPKELAPTTTKESTSTSDKAPTTTPKGAPTTTKPEPAPTTTKPEPAPTTTKPKG | 780 |
| Qy | 738 | TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDKAPTTTKETAPTTTKPEPAPTTTKPEPAPTTPK | 797 |
| Db | 781 | TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDKAPTTTKETAPTTTKPEPAPTTTKPEPAPTTPK | 840 |
| Qy | 798 | KPAPTTPETPTTSEVSTPTTKKEPPTTIHKSPPDESTPELSAEPPTPKALENSPKPGVPT | 857 |
| Db | 841 | KPAPTTPETPTTSEVSTPTTKKEPPTTIHKSPPDESTPELSAEPPTPKALENSPKPGVPT | 900 |
| Qy | 858 | TKTPAATKPEMTTAKOKTTEDRLTPTTETTTAAPKMTKETATTTKTESKITATTTQV | 917 |
| Db | 901 | TKTPAATKPEMTTAKOKTTEDRLTPTTETTTAAPKMTKETATTTKTESKITATTTQV | 960 |
| Qy | 918 | TSITTTQDTTPFKITTLKTTTLAPKVTTTKKITTTTEIMNKPEETAOKPDRAATNSKATTPK | 977 |
| Db | 961 | TSITTTQDTTPFKITTLKTTTLAPKVTTTKKITTTTEIMNKPEETAOKPDRAATNSKATTPK | 1020 |
| Qy | 978 | POKETKAPKPTSTKPKDTPRVRPKKTTTPTPKMTSTMPELNPTDSIAEAMQTTTRPN | 1037 |
| Db | 1021 | POKETKAPKPTSTKPKDTPRVRPKKTTTPTPKMTSTMPELNPTDSIAEAMQTTTRPN | 1080 |
| Qy | 1038 | QTPNSKLVEVNPXSADAGGAGETPHMLLRPHVFMPSVPTDMYDLPRVPMQIILNPMLS | 1097 |
| Db | 1081 | QTPNSKLVEVNPXSADAGGAGETPHMLLRPHVFMPSVPTDMYDLPRVPMQIILNPMLS | 1140 |
| Qy | 1098 | DETNIKNGKPDVGLTTLTNGTLVAFRGHYFWMLSFPSPSPARRITVWGIQSPIDTVFT | 1157 |

Db 200 VETKEITTTNKQSTDCGKEKITSKETSQIEKTSKADLAPTSKVLAAPTPKAETTTKGPA 259
QY 301 LTTPEKPTPTPKPEASTTKPEPTPTTKISAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 360
Db 260 LTTPEKPTPTPKPEASTTKPEPTPTTKISAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 319
QY 361 PAPTTPKEAPTTTKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEPTT 420
Db 320 PAPTTPKEAPTTTKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEPTT 379
QY 421 PKEAPTTKEAPTTTKPEAPTTAPKPEAPTTPKPEAPTTPKPEAPTTTKPEAPTTTKPEPA 480
Db 380 PKEAPTTKEAPTTTKPEAPTTAPKPEAPTTPKPEAPTTPKPEAPTTTKPEAPTTTKPEPA 439
QY 481 PTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 540
Db 440 PTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 499
QY 541 PAPTTPKEAPTTTKPEAPTTAPKPEAPTTPKPEAPTTPKPEAPTTTKPEAPTTTKPEAPTT 600
Db 500 PAPTTPKEAPTTTKPEAPTTAPKPEAPTTPKPEAPTTPKPEAPTTTKPEAPTTTKPEAPTT 559
QY 601 TPEELAPTTPEEPTPTTPEEAPTTPKAAAPNTPEAPTTPKPEAPTTPKPEAPTTTKPEAPTT 660
Db 560 TPEELAPTTPEEPTPTTPEEAPTTPKAAAPNTPEAPTTPKPEAPTTPKPEAPTTTKPEAPTT 619
QY 661 TAPTTPKGTAPTTLKEAPTTTPKPAPELAPTTTKPEPTSTTSKPAEPTTPKGTAPTTTPK 720
Db 620 TAPTTPKGTAPTTLKEAPTTTPKPAPELAPTTTKPEPTSTTSKPAEPTTPKGTAPTTTPK 679
QY 721 EPAPTTPKEAPTTPKGTAPTTLKEAPTTTPKPAPELAPTTTKGPTSTTSKPAEPTTPK 780
Db 680 EPAPTTPKEAPTTPKGTAPTTLKEAPTTTPKPAPELAPTTTKGPTSTTSKPAEPTTPK 739
QY 781 KETAPTTPKEAPTTTPKPAEPTTPKGTAPTTTPKPAPELAPTTTKGPTSTTSKPAEPTTPK 840
Db 740 KETAPTTPKEAPTTTPKPAEPTTPKGTAPTTTPKPAPELAPTTTKGPTSTTSKPAEPTTPK 799
QY 841 PTPKALENSKPEGVPPTTKTAPKPEMTTAKOKTTERDLRTPPETTTAAPTKEPAT 900
Db 800 PTPKALENSKPEGVPPTTKTAPKPEMTTAKOKTTERDLRTPPETTTAAPTKEPAT 859
QY 901 TTEKTTESKIATTTQVSTTTQDTPPKITLTKTTLAPKVTTTKKITTTIEMNKPEE 960
Db 860 TTEKTTESKIATTTQVSTTTQDTPPKITLTKTTLAPKVTTTKKITTTIEMNKPEE 919
QY 961 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPRVRKPTTTPRXTSTMPBLN 1020
Db 920 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPRVRKPTTTPRXTSTMPBLN 979
QY 1021 PTSRIAEMLQTTTRPNQTPNSKLEVNPKSESDAGAGETPHMLLRHVFMPEVTPDMD 1080
Db 980 PTSRIAEMLQTTTRPNQTPNSKLEVNPKSESDAGAGETPHMLLRHVFMPEVTPDMD 1039
QY 1081 YLPRVENOGIILNPMLEDETNIENKGPVDGLTTLRNGTLVAPRGHYFWMLSPFSPSPAR 1140
Db 1040 YLPRVENOGIILNPMLEDETNIENKGPVDGLTTLRNGTLVAPRGHYFWMLSPFSPSPAR 1099
QY 1141 RITEVWGIPSPIDTVTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTQOI 1200
Db 1100 RITEVWGIPSPIDTVTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPKPIFKGFGGLTQOI 1159
QY 1201 VAALSTAKYKNWPEVSYFFKRGGSIOQYIYKQEPVQKCPGRPALNYPYVGMTQVRRRR 1260
Db 1160 VAALSTAKYKNWPEVSYFFKRGGSIOQYIYKQEPVQKCPGRPALNYPYVGMTQVRRRR 1219
QY 1261 FERALGPSOTHIRIQYSFARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1320
Db 1220 FERALGPSOTHIRIQYSFARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1279
QY 1321 GYDYVAFSKDQYVNIIDVPSRTARAITRSGQTLISKWYNCP 1361
Db 1280 GYDYVAFSKDQYVNIIDVPSRTARAITRSGQTLISKWYNCP 1320

RESULT 8
US-07-757-022B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA: US 07/546,114
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-60

Query Match 96.4%; Score 7058.9; DB 4; Length 1320;
Best Local Similarity 97.0%; Pred. No. 5.8e-192;
Matches 1320; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
QY 1 MAWKTLPIYLLILLISVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCDYNQHVMSCCPDF 60
Db 1 MAWKTLPIYLLILLISVFVIQQVSSQ----- 25
QY 61 KRVCVTAELSCKRCRCESFERGERECDDAOCKKYDKCCPDYBSFCAEVHNPSPSSKKAP 120
Db 26 -----ELSCRCRCESFERGERECDDAOCKKYDKCCPDYBSFCAEVHNPSPSSKKAP 79
QY 121 PPSGASQTIKSTTKSPKPPNKKTKVIESEIEITEVKDNKNRTKKKPTPKPPVVDKAP 180
Db 80 PPSGASQTIKSTTKSPKPPNKKTKVIESEIEITEVKDNKNRTKKKPTPKPPVVDKAP 139
QY 181 SGLDNGDFKVTTPDTSITTQHNKVKSTPKITTTAKPINPRESLPPNSDTSKETSILTUNKETT 240

Db 140 SGLDNGDFKVTTPDSTTIOHNVSTSPKTTAKPINPDSLPNSDTSKETSITVKNKETT 199
QY 241 VETKEITTTNKOTSDGKEKTTSAKETSQIEKTSKOLAPTSKVIAKPTPKAETTTKGPA 300
Db 200 VETKEITTTNKOTSDGKEKTTSAKETSQIEKTSKOLAPTSKVIAKPTPKAETTTKGPA 259
QY 301 LTTPEPTTTPKEPASPTTTPKEPTTTTTSAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 360
Db 260 LTTPEPTTTPKEPASPTTTPKEPTTTTTSAPTTTKEPAPTTPKAPTTTKEPAPTTPKE 319
QY 361 PAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 420
Db 320 PAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 379
QY 421 PKBPAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 480
Db 380 PKBPAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 439
QY 481 PTTTTSAPTTTKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 540
Db 440 PTTTTSAPTTTKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 499
QY 541 PAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 600
Db 500 PAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 559
QY 601 TPEELAPTTPEPTTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 660
Db 560 TPEELAPTTPEPTTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 619
QY 661 TAPTTPKGTAPTTTKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 720
Db 620 TAPTTPKGTAPTTTKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 679
QY 721 EPAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 780
Db 680 EPAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 739
QY 781 KETAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 840
Db 740 KETAPTTPKEPAPTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 799
QY 841 PTPKALENSKEPGVPTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 900
Db 800 PTPKALENSKEPGVPTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 859
QY 901 TTEKTTESKITAATTQVTSSTTQDTPPKITTLKTTTLPKVTTKKTTTTEIMNKP 960
Db 860 TTEKTTESKITAATTQVTSSTTQDTPPKITTLKTTTLPKVTTKKTTTTEIMNKP 919
QY 961 TAKPKDRATNSKATTPKAPKPTKAPKPTSTKPKTPMVRKPKTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 1020
Db 920 TAKPKDRATNSKATTPKAPKPTKAPKPTSTKPKTPMVRKPKTTTTPKAPTTTTPKAPTTTTPKAPTTTTPK 979
QY 1021 PTSRIAEMLOTTTRPNQTPNSKLVEVPKSEADAGAGETPHMLLRPHVFMPEVTPDMD 1080
Db 980 PTSRIAEMLOTTTRPNQTPNSKLVEVPKSEADAGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1081 YLPRVFNQGIILINPMLSDETNIENGKVPDGLTTLRNGTLVAFRGHYFWMLSFSPSPSPAR 1140
Db 1040 YLPRVFNQGIILINPMLSDETNIENGKVPDGLTTLRNGTLVAFRGHYFWMLSFSPSPSPAR 1099
QY 1141 RITEVWGIPIPIDTVFTRCNCEGKTFEFDKQSVWRFTNDIKDAGYKPKPIFKGCGGLTGQI 1200
Db 1100 RITEVWGIPIPIDTVFTRCNCEGKTFEFDKQSVWRFTNDIKDAGYKPKPIFKGCGGLTGQI 1159
QY 1201 VAALSTAKYKNWPESVYFFKRGSGIQOYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRRR 1260
Db 1160 VAALSTAKYKNWPESVYFFKRGSGIQOYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRRR 1219
QY 1261 FERAIGPSQTHTRIOYSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPIRKP 1320

Db 1220 FERAIGPSQTHTRIOYSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPIRKP 1279
QY 1321 GVDYVAFSKDOQYINIDVPSRTARAITTRSGQTLISKVWYNCP 1361
Db 1280 GVDYVAFSKDOQYINIDVPSRTARAITTRSGQTLISKVWYNCP 1320

RESULT 9
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-42

Query Match 96.3%; Score 7049; DB 4; Length 1311;
Best Local Similarity 96.3%; Pred. No. 1.le-191;
Matches 1311; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQVYSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVYSSQDLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVTCTAELCKGRCFESFERGECDCDAOCKYDKCCPDYSPCAEVHNPTSPSSKKAP 120
Db 61 KRVTCTAELCKGRCFESFERGECDCDAOCKYDKCCPDYSPCAEVHNPTSPSSKKAP 105

QY 121 PPSGASQTIKSTTKRSKPKNKKTKKVBESBEITEVKNKKNKTKKPTKBPVDEAG 180
Db 106 -----EVKDNKKNKTKKPTKBPVDEAG 130
QY 181 SGLDNGDFKVTTPDTSTTQHNKVSSTPKIITAKBINPRPSLPNNSDTSKETSLTVNKETT 240
Db 131 SGLDNGDFKVTTPDTSTTQHNKVSSTPKIITAKBINPRPSLPNNSDTSKETSLTVNKETT 190
QY 241 VETKETTITNKQTSIDGKEKTSKAKTOSIEKTSKOLAPTSKVLAKPTPKAETTTKGA 300
Db 191 VETKETTITNKQTSIDGKEKTSKAKTOSIEKTSKOLAPTSKVLAKPTPKAETTTKGA 250
QY 301 LTTPEKPTPTPKBPASTTPKEPTTTIKSAPTTPKBPAPTTTKSAPTTKSAPTTPKEPAPTTTKE 360
Db 251 LTTPEKPTPTPKBPASTTPKEPTTTIKSAPTTPKBPAPTTTKSAPTTKSAPTTPKEPAPTTTKE 310
QY 361 PAPTTPKEPAPTTTKAPAPTTTKSAPTTPKBPAPTTTPKKPAPTTTPKEPAPTTTPKEPTPTT 420
Db 311 PAPTTPKEPAPTTTKAPAPTTTKSAPTTPKBPAPTTTPKKPAPTTTPKEPAPTTTPKEPTPTT 370
QY 421 PKBPAPTTKEPAPTTPKBPAPTAAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 480
Db 371 PKBPAPTTKEPAPTTPKBPAPTAAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 430
QY 481 PTTTKSAPTTTKAPAPTTTKSAPTTPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 540
Db 431 PTTTKSAPTTTKAPAPTTTKSAPTTPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 490
QY 541 PAPTTPKEPAPTTTKKAPAPTAAPKAPAPTTTPKETAPTTPKLTPPTPEKAPAPTTPEKAPPT 600
Db 491 PAPTTPKEPAPTTTKKAPAPTAAPKAPAPTTTPKETAPTTPKLTPPTPEKAPAPTTPEKAPPT 550
QY 601 TPBELAPTTPEPTPTTPEEPAPTTPKAAANWPKBPAPTTTPKEPAPTTTPKEPAPTTTPKE 660
Db 551 TPBELAPTTPEPTPTTPEEPAPTTPKAAANWPKBPAPTTTPKEPAPTTTPKEPAPTTTPKE 610
QY 661 TAPPTPKGTAPTTLKEPAPTTPKAPKELAPTTTKETPTSTSDKAPPTPKGTAPTTTPK 720
Db 611 TAPPTPKGTAPTTLKEPAPTTPKAPKELAPTTTKETPTSTSDKAPPTPKGTAPTTTPK 670
QY 721 EPAPTTPKAPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKGTPTSTSDKAPPTTP 780
Db 671 EPAPTTPKAPAPTTPKGTAPTTLKEPAPTTPKAPKELAPTTTKGTPTSTSDKAPPTTP 730
QY 781 KETAPTTKEPAPTTPKAPAPTTTPPTPTTSEVSTPTTKEPTTIHKSPESTBELSAE 840
Db 731 KETAPTTKEPAPTTPKAPAPTTTPPTPTTSEVSTPTTKEPTTIHKSPESTBELSAE 790
QY 841 PTEKALENSPKEPVPPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETAT 900
Db 791 PTEKALENSPKEPVPPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETAT 850
QY 901 TTEKTTESKITATTTQVTSITTTQDTPPKITTLKTTTLAPKVTTKTITTEIMNKPEE 960
Db 851 TTEKTTESKITATTTQVTSITTTQDTPPKITTLKTTTLAPKVTTKTITTEIMNKPEE 910
QY 961 TAKPKDRATNSKATTPKQKPKAKKPTSTTKPKTMRVRKPTTTPRKMSTMPBELN 1020
Db 911 TAKPKDRATNSKATTPKQKPKAKKPTSTTKPKTMRVRKPTTTPRKMSTMPBELN 970
QY 1021 PTSRIAEMLOTTTTRNQTPNSKLVEVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMD 1080
Db 971 PTSRIAEMLOTTTTRNQTPNSKLVEVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMD 1030
QY 1081 YLPRVNPQGIINPMLSDETNIENCGKPVGLTTLRNGTLVAPRGHYFWMLSFPSPSPAR 1140
Db 1031 YLPRVNPQGIINPMLSDETNIENCGKPVGLTTLRNGTLVAPRGHYFWMLSFPSPSPAR 1090
QY 1141 RITEVWGIESPIDTIVTRCNCEKTFPFKDSQVWRFTNDIKDAGVPKPIFKFGGLTGQI 1200
Db 1091 RITEVWGIESPIDTIVTRCNCEKTFPFKDSQVWRFTNDIKDAGVPKPIFKFGGLTGQI 1150
QY 1201 VAALSTAKYNWPESVYFFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRR 1260

Db 1151 VAALSTAKYNWPESVYFFKRGGSIOQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRR 1210
QY 1261 FERAIGPSQTHTRIQYSPARLAYODKGVILHNEVKVSIILWRGLPNVVTSAISLENIRKPD 1320
Db 1211 FERAIGPSQTHTRIQYSPARLAYODKGVILHNEVKVSIILWRGLPNVVTSAISLENIRKPD 1270
QY 1321 GYDYAFSKQOYNNIDVPSRTARAITTRSGQTLISKWYNCP 1361
Db 1271 GYDYAFSKQOYNNIDVPSRTARAITTRSGQTLISKWYNCP 1311
RESULT 10
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geerz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-52
Query Match 96.2%; Score 7044.6; DB 4; Length 1363;
Best Local Similarity 94.0%; Pred. No. 1.5e-191;
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;
QY 1 MAWKTLPIYLLILLLSVFVIQVSSQDLSSCAGCGEGYSRDATCNCDYNQHYMECCPDF 60
Db 1 MAWKTLPIYLLILLLSVFVIQVSSQ----- 25

QY 61 KRVTAEISCKGRCFESFERGECDDAOCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 26 -----ELSCKGRCFESFERGECDDAOCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITE-----156
DB 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEEHSVSENQBSSSSSSSSSSSTIw 139
QY 157 -----VKONKNRTKKKPTPKPPVVDAGSGLNDGPKVTTPTST 197
DB 140 KIKSSKNSAANRELQKKLVKONKNRTKKKPTPKPPVVDAGSGLNDGPKVTTPTST 199
QY 198 TQHNKYSTSKITTTAKPINRPSLPNSDTSKNSLTVNKETTVEKETTNNKQTSDDG 257
DB 200 TQHNKYSTSKITTTAKPINRPSLPNSDTSKNSLTVNKETTVEKETTNNKQTSDDG 259
QY 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEBPAS 317
DB 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEBPAS 319
QY 318 TTPKEPTPTTIKSAPTTPKEBPATTTTSAPTTPKEBPATTTTKEBPATTTTKEP 377
DB 320 TTPKEPTPTTIKSAPTTPKEBPATTTTSAPTTPKEBPATTTTKEBPATTTTKEP 379
QY 378 APITTTKSAPTTPKEBPATTPKPPAPTTTPKEBPATTPKEPTTTPKEBPATTPK 437
DB 380 APITTTKSAPTTPKEBPATTPKPPAPTTTPKEBPATTPKEPTTTPKEBPATTPK 439
QY 438 EPAPTAPKKAPPTTPKEBPATTPKEBPATTTTKEPSPTTPKEBPATTTTKEBPAT 497
DB 440 EPAPTAPKKAPPTTPKEBPATTPKEBPATTTTKEPSPTTPKEBPATTTTKEBPAT 499
QY 498 TTKSAPTTPKEPSPTTTPKEBPATTPKEBPATTPKPPAPTTTPKEBPATTTTKEP 557
DB 500 TTKSAPTTPKEPSPTTTPKEBPATTPKEBPATTPKPPAPTTTPKEBPATTTTKEP 559
QY 558 APTAPKEPATTPKETAPTTPKKLTPTTPKELAPTTPEKAPATTPPELAPTTPEPTPTT 617
DB 560 APTAPKEPATTPKETAPTTPKKLTPTTPKELAPTTPEKAPATTPPELAPTTPEPTPTT 619
QY 618 PEPAPPTPKAAANPTPKAPATTPKEBPATTPKEBPATTPKETAPTTPKGTAPTTLKEP 677
DB 620 PEPAPPTPKAAANPTPKAPATTPKEBPATTPKEBPATTPKETAPTTPKGTAPTTLKEP 679
QY 678 APTTPKKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTPKEBPATTPKGBAPTTPKG 737
DB 680 APTTPKKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTPKEBPATTPKGBAPTTPKG 739
QY 738 TAPTTLKEPATTPPKPAPKELAPTTTKGPTSTSDKAPATTPKETAPTTPKEBPATTPK 797
DB 740 TAPTTLKEPATTPPKPAPKELAPTTTKGPTSTSDKAPATTPKETAPTTPKEBPATTPK 799
QY 798 KPAPTTPETPPPTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSPKGCVPT 857
DB 800 KPAPTTPETPPPTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSPKGCVPT 859
QY 858 TKTPAATKPMITTAKDKITTERDLRTTPETTTAAPKMTKATATTETKTESKITATTQV 917
DB 860 TKTPAATKPMITTAKDKITTERDLRTTPETTTAAPKMTKATATTETKTESKITATTQV 919
QY 918 TSTTTQDTPFKITTLTKTTLAPKVTITTKTITTEIMNKPEETAKPKDRATNSKATTPK 977
DB 920 TSTTTQDTPFKITTLTKTTLAPKVTITTKTITTEIMNKPEETAKPKDRATNSKATTPK 979
QY 978 PQKPTKAPKKPTSTKPKTMVRVKPXTTTPRKMTSTMPELNPTSGRIAEAMLTQTTTRN 1037
DB 980 PQKPTKAPKKPTSTKPKTMVRVKPXTTTPRKMTSTMPELNPTSGRIAEAMLTQTTTRN 1039
QY 1038 QTNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMDEVTPDMOYLPRVNPQGIINPMLS 1097
DB 1040 QTNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMDEVTPDMOYLPRVNPQGIINPMLS 1099

QY 1098 DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITTEVWGIPSPIDTIVFT 1157
DB 1100 DETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITTEVWGIPSPIDTIVFT 1159
QY 1158 RCNCEGXTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1217
DB 1160 RCNCEGXTFFFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1219
QY 1218 FFKRGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERAIGPSQTHIRIQY 1277
DB 1220 FFKRGSGSIQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFERAIGPSQTHIRIQY 1279
QY 1278 SPARLAYODKGVHLNVEKVSILMRGLPNVVTSAISLEINIRKPDGYDYAFSKDQYVYNDV 1337
DB 1280 SPARLAYODKGVHLNVEKVSILMRGLPNVVTSAISLEINIRKPDGYDYAFSKDQYVYNDV 1339
QY 1338 PSRTARAITTRSGQTLSKVMWYVNC 1361
DB 1340 PSRTARAITTRSGQTLSKVMWYVNC 1363
RESULT 11
US-10-164-595-58
; Sequence 58, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 58
; LENGTH: 1320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-58
Query Match 96.2%; Score 7043.9; DB 4; Length 1320;
Best Local Similarity 96.8%; Pred. No. 1.5e-191;
Matches 1317; Conservative 0; Mismatches 3; Indels 41; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIOQVSSQDLSSCAGCGEGYSRDATCNCYNCQHMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIOQVSSQ-----25
QY 61 KRVTAEISCKGRCFESFERGECDDAOCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
DB 26 -----ELSCKGRCFESFERGECDDAOCKKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
QY 121 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEVYKONKNRTKKKPTPKPPVVDAG 180
DB 80 PPSGASQTIKSTTKRSPKPNKKTKKVIIESEITEVYKONKNRTKKKPTPKPPVVDAG 139
QY 181 SGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT 240
DB 140 SGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT 199
QY 241 VETKETTTNNKOTSDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGA 300
DB 200 VETKETTTNNKOTSDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGA 259
QY 301 LTTPEKPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEBPATTTTSAPTTPKEBPATTTTKE 360
DB 260 LTTPEKPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEBPATTTTSAPTTPKEBPATTTTKE 319
QY 361 PAPTTPKEBPATTTTKEBPATTTTSAPTTPKEBPATTPPKAPATTPKEBPATTPKEPTPTT 420
DB 320 PAPTTPKEBPATTTTKEBPATTTTSAPTTPKEBPATTPPKAPATTPKEBPATTPKEPTPTT 379
QY 421 PKEPAPTTPKEBPATTPKEBPATTPKAPATTPKEBPATTPKEBPATTTTKEBPATTTTKEPA 480

QY 301 LITPKBPTTTPKEPASTTTPKEPTTTPKISAPTTTPKEPAPTTPKSAPTTTPKEBAPTTPKE 360
DB 253 LITPKBPTTTPKEPASTTTPKEPTTTPKISAPTTTPKEPAPTTPKSAPTTTPKEBAPTTPKE 312
QY 361 PAPTTPKEPAPTTPKBPATTTKASPTTTPKEPAPTTPKBPATTPKEPAPTTPKEBAPTTPKE 420
DB 313 PAPTTPKEPAPTTPKBPATTTKASPTTTPKEPAPTTPKBPATTPKEPAPTTPKEBAPTTPKE 372
QY 421 PKBPAPTTPKEPAPTTPKEPAPTTPKBPATTPKEPAPTTPKEPAPTTPKEBAPTTPKEBPA 480
DB 373 PKBPAPTTPKEPAPTTPKEPAPTTPKBPATTPKEPAPTTPKEPAPTTPKEBAPTTPKEBPA 432
QY 481 PTTTTSAPTTTPKEPAPTTPKASPTTTPKEPAPTTPKBPATTPKEPAPTTPKEBAPTTPKE 540
DB 433 PTTTTSAPTTTPKEPAPTTPKASPTTTPKEPAPTTPKBPATTPKEPAPTTPKEBAPTTPKE 492
QY 541 PAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 600
DB 493 PAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 552
QY 601 TPELAPTTPKEPAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPTTPKE 660
DB 553 TPELAPTTPKEPAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPTTPKE 612
QY 661 TAPTTPKGTAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 720
DB 613 TAPTTPKGTAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 672
QY 721 EPAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 780
DB 673 EPAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 732
QY 781 KETAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 840
DB 733 KETAPTTPKEPAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 792
QY 841 PTPKALNSPKBPVPTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 900
DB 793 PTPKALNSPKBPVPTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 852
QY 901 TTEKTTESKIATTTTQVSTTTTQDTTPFKITLTKITTLAPKVTTTKITTTTEIMNKPEE 960
DB 853 TTEKTTESKIATTTTQVSTTTTQDTTPFKITLTKITTLAPKVTTTKITTTTEIMNKPEE 912
QY 961 TAKPKDRATNSKATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 1020
DB 913 TAKPKDRATNSKATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 972
QY 1021 PTSRIAEAMLQTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 1080
DB 973 PTSRIAEAMLQTTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKEBAPT 1032
QY 1081 YLPRVFNQGIITINPMSDETNI CNCKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1140
DB 1033 YLPRVFNQGIITINPMSDETNI CNCKPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1092
QY 1141 RITEVWGIPSPIDTFTTRCNCEGKTFPFKDSQYWFNDIKDAGYKPKIFKFGGLGTQI 1200
DB 1093 RITEVWGIPSPIDTFTTRCNCEGKTFPFKDSQYWFNDIKDAGYKPKIFKFGGLGTQI 1152
QY 1201 VAALSTAKYKNWPFESVYFPRGSGIQOYIYKQEPVQKCPGRPALNPYVGMTOVRRRR 1260
DB 1153 VAALSTAKYKNWPFESVYFPRGSGIQOYIYKQEPVQKCPGRPALNPYVGMTOVRRRR 1212
QY 1261 FERAIGPSQTHIRIQYSPARLAYQDKVGLHNEVKVSIILWRGLPNVVTSAISLPIRKPDP 1320
DB 1213 FERAIGPSQTHIRIQYSPARLAYQDKVGLHNEVKVSIILWRGLPNVVTSAISLPIRKPDP 1272
QY 1321 GYDYAFSKDQYINIDVPRTARAITTRSGQTLISKWYNCP 1361
DB 1273 GYDYAFSKDQYINIDVPRTARAITTRSGQTLISKWYNCP 1313

RESULT 14
US-07-757-022B-44
; Sequence 44, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757.022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseerr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; LENGTH: 1270 amino acids
; SEQUENCE CHARACTERISTICS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-44
Query Match 92.7%; Score 6784.9; DB 4; Length 1270;
Best Local Similarity 93.3%; Pred. No. 3.1e-184;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSCACGCGYSDATCNDYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KRVCCTAELSCGRCPESFERGREGCDCAQCKYDKCCPDYEFCAEHNPTSPSSKKAP 120
DB 26 -----ELSCGRCPESFERGREGCDCAQCKYDKCCPDYEFCA----- 64
QY 121 PPSGASQTIKSTTKSPKPPNKKTKVIESEITEVKNKNRTPKPTPKPPVVDG 180
DB 65 -----EVDKNKNRTPKPTPKPPVVDG 89
QY 181 SGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPNRPSLPNSDTSKETSITVKNKETT 240
DB 90 SGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPNRPSLPNSDTSKETSITVKNKETT 149

QY 241 VETKETTTTNKQTSIDGKEKTTSAKETSADLAPTSKVLAKPTPKAETTTKGPA 300
Db 150 VETKETTTTNKQTSIDGKEKTTSAKETSADLAPTSKVLAKPTPKAETTTKGPA 209
QY 301 LITPKPTTTPKEPASTTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKE 360
Db 210 LITPKPTTTPKEPASTTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKE 269
QY 361 PAPTTPKEPAPTTTPKEPASTTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKE 420
Db 270 PAPTTPKEPAPTTTPKEPASTTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKE 329
QY 421 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 480
Db 330 PKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 389
QY 481 PTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 540
Db 390 PTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 449
QY 541 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 600
Db 450 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 509
QY 601 TPEELAPTTTPEPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 660
Db 510 TPEELAPTTTPEPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 569
QY 661 TAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 720
Db 570 TAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 629
QY 721 BPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 780
Db 630 BPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 689
QY 781 KETAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 840
Db 690 KETAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 749
QY 841 PTPKALNSPKPEPGVPTTKTAAPKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAT 900
Db 750 PTPKALNSPKPEPGVPTTKTAAPKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETAT 809
QY 901 TTEKTESKIATTTQVTSITTTQDTPPKITLTKTTTLAPKVTITTKIITTEIMNKPEE 960
Db 810 TTEKTESKIATTTQVTSITTTQDTPPKITLTKTTTLAPKVTITTKIITTEIMNKPEE 869
QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMRVRKPKTTPTRKWTSTMPBLN 1020
Db 870 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMRVRKPKTTPTRKWTSTMPBLN 929
QY 1021 PTSIAEAMLOTTTRPNQTPNSKLVVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 1080
Db 930 PTSIAEAMLOTTTRPNQTPNSKLVVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMD 989
QY 1081 YLPRVPMQGIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPAR 1140
Db 990 YLPRVPMQGIILNPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSPAR 1049
QY 1141 RITEVWGIPIPIDVTRFRCNCEGKTFKPKSQYWRFTNDIKDAGVPKPIFKGFGGLTGOI 1200
Db 1050 RITEVWGIPIPIDVTRFRCNCEGKTFKPKSQYWRFTNDIKDAGVPKPIFKGFGGLTGOI 1109
QY 1201 VAALSTAKYKNWPSVVFVKFGGSIQQYIKQEPVQKPCGRRPALNYPVVGEMTQVRRR 1260
Db 1110 VAALSTAKYKNWPSVVFVKFGGSIQQYIKQEPVQKPCGRRPALNYPVVGEMTQVRRR 1169
QY 1261 FERAIGPSQTHIRIQSPARLAYODKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1320
Db 1170 FERAIGPSQTHIRIQSPARLAYODKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKPD 1229

QY 1321 GYDYAFSKDOYNNIDVPSRTARAITTRSGQTLISKVWYNCP 1361
Db 1230 GYDYAFSKDOYNNIDVPSRTARAITTRSGQTLISKVWYNCP 1270
RESULT 15
US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseer, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-104
Query Match 80.1%; Score 5867.7; DB 4; Length 1140;
Best Local Similarity 96.2%; Pred. No. 2.5e-158;
Matches 1097; Conservative 0; Mismatches 0; Indels 43; Gaps 1;
QY 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNQCHYMECCPDF 60
Db 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNQCHYMECCPDF 60
QY 61 KRVCYTAELSCKGRCFESFERGRCDCDAQCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120
Db 61 KRVCYTAELSCKGRCFESFERGRCDCDAQCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120
QY 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEITE----- 156
Db 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEITE----- 156

Db 121 PPGASQTIKSTTRSPKPPNKKTKKVIIESEBITEHSHSVSEHQESSSSSSSSSSSTIWM 180
QY 157 -----VKDNKNRTKKKPTKPPVVDVAGSGLONGDFKVITPDST 197
Db 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTKPPVVDVAGSGLONGDFKVITPDST 240
QY 198 TOHNKVSTSPKITTAKPILNPPSLPPNSDTSKETSILVKNKETTIVETKETTNNKQTSIDG 257
Db 241 TOHNKVSTSPKITTAKPILNPPSLPPNSDTSKETSILVKNKETTIVETKETTNNKQTSIDG 300
QY 258 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAEITTKGPAITTKPEPTTKPEPAS 317
Db 301 KEKTTSAKETOSIEKTSKADLAPTSKVLAKPTPKAEITTKGPAITTKPEPTTKPEPAS 360
QY 318 TTPKEPTTTIKSAFTTTPKEPAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEP 377
Db 361 TTPKEPTTTIKSAFTTTPKEPAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEP 420
QY 378 APTTTKSAPTTTPKEPAPTTTPKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 437
Db 421 APTTTKSAPTTTPKEPAPTTTPKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480
QY 438 EPAPTAPKKPAPTTPKEPAPTTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPT 497
Db 481 EPAPTAPKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPT 540
QY 498 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKKPAPTTPKEPAPTTTPKEPAPTTTPKKP 557
Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKKPAPTTPKEPAPTTTPKEPAPTTTPKKP 600
QY 558 APTAKPEPAPTTTPKEPAPTTTPKKLPTTPKEPAPTTTPKEPAPTTTPPELAPTTPPEPTPTT 617
Db 601 APTAKPEPAPTTTPKEPAPTTTPKKLPTTPKEPAPTTTPKEPAPTTTPPELAPTTPPEPTPTT 660
QY 618 PREPAPTTPKAAAPNTPKPAPTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGAPTTLKEP 677
Db 661 PREPAPTTPKAAAPNTPKPAPTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGAPTTLKEP 720
QY 678 APTTPKKPAPKELAPTTTKPEPTSTSDKAPATTPKGAPTTPKEPAPTTTPKEPAPTTTPKG 737
Db 721 APTTPKKPAPKELAPTTTKPEPTSTSDKAPATTPKGAPTTPKEPAPTTTPKEPAPTTTPKG 780
QY 738 TAPTTILKEPAPTTPKKPAPKELAPTTTKGPTSTSDKAPATTPKETAPTTPKEPAPTTTPK 797
Db 781 TAPTTILKEPAPTTPKKPAPKELAPTTTKGPTSTSDKAPATTPKETAPTTPKEPAPTTTPK 840
QY 798 KDAPTTPETPPPTTSVSTPTTTKEPTTIHKSPDESTPELSAFTPKALENSPKBPGVPT 857
Db 841 KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSPDESTPELSAFTPKALENSPKBPGVPT 900
QY 858 TKTPAAATKPEMTTAKDKTERDLRTPPTTAAAPKOTKETATTTTEKTTESKITATTQV 917
Db 901 TKTPAAATKPEMTTAKDKTERDLRTPPTTAAAPKOTKETATTTTEKTTESKITATTQV 960
QY 918 TSTTTQDTPPKITTLKTTTLAPKVTTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 977
Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 1020
QY 978 FOKPTKAPKKPTSTKKPKTTPVRKPKTTTPPKMTSTMPELNPTSRIAEAMLQTTTREN 1037
Db 1021 FOKPTKAPKKPTSTKKPKTTPVRKPKTTTPPKMTSTMPELNPTSRIAEAMLQTTTREN 1080
QY 1038 QTPNSKLIVNPKSEBAGGEGTTPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1097
Db 1081 QTPNSKLIVNPKSEBAGGEGTTPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1140

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 138.466 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-G
Perfect score: 7323
Sequence: 1 MAWKTLPIVLLLLSVFVIQ.....ARAITRSGQTLKVMWVNC 1361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------------------------------|
| 1 | 7323 | 100.0 | 1361 | 13 | US-10-124-557-40 Sequence 40, Appl |
| 2 | 7308.7 | 99.8 | 1404 | 9 | US-09-802-207-30 Sequence 30, Appl |
| 3 | 7308.7 | 99.8 | 1404 | 11 | US-09-897-188-1 Sequence 1, Appli |
| 4 | 7308.7 | 99.8 | 1404 | 13 | US-10-124-557-2 Sequence 2, Appli |
| 5 | 7308.7 | 99.8 | 1404 | 13 | US-10-124-557-62 Sequence 62, Appl |
| 6 | 7084.3 | 96.7 | 1354 | 13 | US-10-124-557-48 Sequence 48, Appl |
| 7 | 7058.9 | 96.4 | 1320 | 13 | US-10-124-557-46 Sequence 46, Appl |
| 8 | 7058.9 | 96.4 | 1320 | 13 | US-10-124-557-60 Sequence 60, Appl |
| 9 | 7049 | 96.3 | 1311 | 13 | US-10-124-557-42 Sequence 42, Appl |
| 10 | 7044.6 | 96.2 | 1363 | 13 | US-10-124-557-52 Sequence 52, Appl |
| 11 | 6836.3 | 93.4 | 1314 | 13 | US-10-124-557-50 Sequence 50, Appl |
| 12 | 6820.2 | 93.1 | 1313 | 13 | US-10-124-557-142 Sequence 142, App |
| 13 | 6784.9 | 92.7 | 1270 | 13 | US-10-124-557-44 Sequence 44, Appl |
| 14 | 5867.7 | 80.1 | 1140 | 13 | US-10-124-557-104 Sequence 104, App |

| | | | | | | |
|----|--------|------|-------|----|---------------------|-------------------|
| 15 | 5485.7 | 74.9 | 1038 | 13 | US-10-124-557-74 | Sequence 74, Appl |
| 16 | 5472 | 74.7 | 1022 | 13 | US-10-124-557-84 | Sequence 84, Appl |
| 17 | 5379.2 | 73.5 | 1049 | 13 | US-10-124-557-58 | Sequence 58, Appl |
| 18 | 5011 | 68.4 | 941 | 13 | US-10-124-557-14 | Sequence 14, Appl |
| 19 | 3944.5 | 53.9 | 792 | 9 | US-09-802-207-27 | Sequence 27, Appl |
| 20 | 2850.9 | 38.9 | 538 | 14 | US-10-038-694-3 | Sequence 3, Appli |
| 21 | 2246.2 | 30.7 | 463 | 13 | US-10-124-557-54 | Sequence 54, Appl |
| 22 | 1991.2 | 27.2 | 423 | 13 | US-10-124-557-66 | Sequence 66, Appl |
| 23 | 1982.1 | 27.1 | 422 | 13 | US-10-124-557-68 | Sequence 68, Appl |
| 24 | 1723.1 | 23.5 | 372 | 13 | US-10-124-557-64 | Sequence 64, Appl |
| 25 | 1714.7 | 23.4 | 401 | 9 | US-09-802-207-29 | Sequence 4, Appli |
| 26 | 1666.9 | 21.4 | 292 | 16 | US-10-468-910-4 | Sequence 1068, Ap |
| 27 | 1413.6 | 19.3 | 5179 | 9 | US-09-922-217-1068 | Sequence 1068, Ap |
| 28 | 1413.6 | 19.3 | 5179 | 9 | US-09-833-263-1068 | Sequence 1068, Ap |
| 29 | 1413.6 | 19.3 | 5179 | 13 | US-10-025-380-1068 | Sequence 1068, Ap |
| 30 | 1413.6 | 19.3 | 5179 | 16 | US-10-734-564-121 | Sequence 121, App |
| 31 | 1295.5 | 17.7 | 296 | 13 | US-10-124-557-70 | Sequence 70, Appl |
| 32 | 1214.6 | 16.6 | 237 | 13 | US-10-124-557-72 | Sequence 72, Appl |
| 33 | 1143 | 15.6 | 207 | 13 | US-10-124-557-116 | Sequence 116, App |
| 34 | 1143 | 15.6 | 207 | 13 | US-10-124-557-136 | Sequence 136, App |
| 35 | 1025.9 | 14.0 | 185 | 16 | US-10-468-910-2 | Sequence 2, Appli |
| 36 | 1010 | 13.8 | 188 | 14 | US-10-038-694-2 | Sequence 2, Appli |
| 37 | 992.5 | 13.6 | 1460 | 14 | US-10-295-027-428 | Sequence 428, App |
| 38 | 983.3 | 13.4 | 3507 | 14 | US-10-369-493-5784 | Sequence 5784, Ap |
| 39 | 971.4 | 13.3 | 6642 | 14 | US-10-369-493-5013 | Sequence 5013, Ap |
| 40 | 971 | 13.3 | 1367 | 9 | US-09-801-368-108 | Sequence 108, App |
| 41 | 970.7 | 13.3 | 220 | 13 | US-10-124-557-96 | Sequence 96, Appl |
| 42 | 957.9 | 13.1 | 19723 | 15 | US-10-084-846A-5 | Sequence 5, Appli |
| 43 | 956 | 13.1 | 231 | 13 | US-10-124-557-30 | Sequence 30, Appl |
| 44 | 953.9 | 13.0 | 1325 | 9 | US-09-864-761-35612 | Sequence 35612, A |
| 45 | 922.8 | 12.6 | 5935 | 14 | US-10-243-243A-8 | Sequence 8, Appli |

ALIGNMENTS

RESULT 1

US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine C.
; Jacobs, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseri, Luan
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: G1 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match 100.0%; Score 7323; DB 13; Length 1361;
Best Local Similarity 100.0%; Pred. No. 1.7e-169;
Matches 1361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLSVFIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

QY 61 KRVCIAELSCKRCFESFERGECDDCAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120
Db 61 KRVCIAELSCKRCFESFERGECDDCAOCKYDKCCPDYBSCAEVHNPTSPSSKKAP 120

QY 121 PPSGASQTIKSTTRSPKPPNKKTKVIESEITEVKDNKKNRKTKKPTKPPVWDEAG 180
Db 121 PPSGASQTIKSTTRSPKPPNKKTKVIESEITEVKDNKKNRKTKKPTKPPVWDEAG 180

QY 181 SGLDNGDFKVTPTDSTTOHKNVSTSPKIITAKDINRPSPSLPNSDTSKETSIVNKEIT 240
Db 181 SGLDNGDFKVTPTDSTTOHKNVSTSPKIITAKDINRPSPSLPNSDTSKETSIVNKEIT 240

QY 241 VETKETITTNKQSTSDGKEKTTSAKETQSTKTSKOLAPTSKVLAKPTPKAETTTKGA 300
Db 241 VETKETITTNKQSTSDGKEKTTSAKETQSTKTSKOLAPTSKVLAKPTPKAETTTKGA 300

QY 301 LTTPEKPTTPPKBPASTTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKE 360
Db 301 LTTPEKPTTPPKBPASTTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKE 360

QY 361 PAPTPKEPAPTTTKBPAPTTTKSAPTTKEPAPTTPKKAPTTKBPAPTTTKEPAPTTT 420
Db 361 PAPTPKEPAPTTTKBPAPTTTKSAPTTKEPAPTTPKKAPTTKBPAPTTTKEPAPTTT 420

QY 421 PKEPAPTTKBPAPTTKBPAPTTAPKBPAPTTKBPAPTTKBPAPTTTKEPAPTTTKEPAP 480
Db 421 PKEPAPTTKBPAPTTKBPAPTTAPKBPAPTTKBPAPTTKBPAPTTTKEPAPTTTKEPAP 480

QY 481 PTTTKSAPTTTKBPAPTTKSAPTTKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540
Db 481 PTTTKSAPTTTKBPAPTTKSAPTTKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540

QY 541 PAPTPKEPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTTKEPAPTTTKEPAPTT 600
Db 541 PAPTPKEPAPTTTKBPAPTTKBPAPTTKBPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600

QY 601 TPELAPTTPEEPTTPEEAPTTPKAAAPNTKPEPAPTTKPEPAPTTTKEPAPTTTKEPAP 660
Db 601 TPELAPTTPEEPTTPEEAPTTPKAAAPNTKPEPAPTTKPEPAPTTTKEPAPTTTKEPAP 660

QY 661 TAPTPKGTAPTTTKBPAPTTPKKAPKELAPTTTKEPTSTSDKPAPTTTPKGPAPTTTPK 720
Db 661 TAPTPKGTAPTTTKBPAPTTPKKAPKELAPTTTKEPTSTSDKPAPTTTPKGPAPTTTPK 720

QY 721 EPAPTTTKEPAPTTTPKGPAPTTTLKBPAPTTPKKAPKELAPTTTKEPTSTSDKPAPTT 780
Db 721 EPAPTTTKEPAPTTTPKGPAPTTTLKBPAPTTPKKAPKELAPTTTKEPTSTSDKPAPTT 780

QY 781 KETAPTTKBPAPTTTPKGPAPTTTTPETPPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAE 840

Db 781 KETAPTTKBPAPTTTPKGPAPTTTTPETPPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAE 840

QY 841 PTPKALENSPKPEPGVPTTKTTPAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETAT 900
Db 841 PTPKALENSPKPEPGVPTTKTTPAATKPEMTTTAKDKTTERDLRTTPTTAAAPKMTKETAT 900

QY 901 TTEKTTESKITATTTQVTSITTTQDTPPKITTLTKTTLAPKVTTKTITTTTEIMNKPEE 960
Db 901 TTEKTTESKITATTTQVTSITTTQDTPPKITTLTKTTLAPKVTTKTITTTTEIMNKPEE 960

QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTPTPRKMWTSMPELN 1020
Db 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKTWPVRVKPKTTPTPRKMWTSMPELN 1020

QY 1021 PTSRIAEMLQTTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPMD 1080
Db 1021 PTSRIAEMLQTTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPMD 1080

QY 1081 YLPRVFNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1140
Db 1081 YLPRVFNQGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1140

QY 1141 RITEVWGIPIPDVFTFRCNCEGKTFEKKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGI 1200
Db 1141 RITEVWGIPIPDVFTFRCNCEGKTFEKKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGI 1200

QY 1201 VAALSTAKYKNWPESVYFFKRGSGIQYIYKQEPVQKCPGRRPALNPVYVGMTQVRRRR 1260
Db 1201 VAALSTAKYKNWPESVYFFKRGSGIQYIYKQEPVQKCPGRRPALNPVYVGMTQVRRRR 1260

QY 1261 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWGLPNVVTSAISLENIRKPD 1320
Db 1261 FERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWGLPNVVTSAISLENIRKPD 1320

QY 1321 GYDYAFAKQYQYNNIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 1321 GYDYAFAKQYQYNNIDVPSRTARAITTRSGOTLSKWYNCP 1361

RESULT 2

US-09-802-207-30
; Sequence 30, Application US/09802207
; Publication No. US20020086824A1
; GENERAL INFORMATION:
; APPLICANT: Warman, Matthew
; APPLICANT: Carpten, John
; APPLICANT: Trent, Jeffrey
; APPLICANT: Marcelino, Jose
; TITLE OF INVENTION: Novel Methods and Reagents for the Treatment of Osteoarthritis
; FILE REFERENCE: Case-06212
; CURRENT APPLICATION NUMBER: US/09/802,207
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 09/619,175
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,328
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-207-30

Query Match 99.8%; Score 7308.7; DB 9; Length 1404;
Best Local Similarity 96.9%; Pred. No. 3.9e-169;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

QY 258 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAEITTKGPAITTPKEPTTTPKEPAS 317
Db 301 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAEITTKGPAITTPKEPTTTPKEPAS 360
QY 318 TTPKEPTTTIKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 377
Db 361 TTPKEPTTTIKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 420
QY 378 APTTTKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTPK 437
Db 421 APTTTKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTPK 480
QY 438 EPAPTAPKAPPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTT 497
Db 481 EPAPTAPKAPPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTT 540
QY 498 TTKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 557
Db 541 TTKSAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 600
QY 558 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKLAPTTPEELAPTTPPEPTTT 617
Db 601 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPEKLAPTTPEELAPTTPPEPTTT 660
QY 618 PEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 677
Db 661 PEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 720
QY 678 APTPKKPAKELAPTTPKEPTSTTSOKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG 737
Db 721 APTPKKPAKELAPTTPKEPTSTTSOKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG 780
QY 738 TAPTTLKEPAPTTPKPAKELAPTTPKGTSTSDKPAPTTPKETAAPTTPKEPAPTTPK 797
Db 781 TAPTTLKEPAPTTPKPAKELAPTTPKGTSTSDKPAPTTPKETAAPTTPKEPAPTTPK 840
QY 798 KPAPTTPETPTTSEVSTTTTKEPTTIHKSDESTPELSAETTPKALENSKPEGVPT 857
Db 841 KPAPTTPETPTTSEVSTTTTKEPTTIHKSDESTPELSAETTPKALENSKPEGVPT 900
QY 858 TKTPAATKEMTTAKDKTTERDLRTTPEPTTAAAPKMTKETAATTEKTESKITATTQV 917
Db 901 TKTPAATKEMTTAKDKTTERDLRTTPEPTTAAAPKMTKETAATTEKTESKITATTQV 960
QY 918 TSTTTQDTPFKITTLTKTTLAPKVTTTKTITTEIMNKPEETAKPKDRAATNSKATTPK 977
Db 961 TSTTTQDTPFKITTLTKTTLAPKVTTTKTITTEIMNKPEETAKPKDRAATNSKATTPK 1020
QY 978 PQKPTKAPKPTSTKXPKTMVRVKPTTTPRKMTSTMPELNPTSGRIAEAMLQTTTRPN 1037
Db 1021 PQKPTKAPKPTSTKXPKTMVRVKPTTTPRKMTSTMPELNPTSGRIAEAMLQTTTRPN 1080
QY 1038 QTPNSKLVEVPKSEDAGGAGETPHMLLRPHVFMPEVTPDMYLLPRVNOGIIINPMLS 1097
Db 1081 QTPNSKLVEVPKSEDAGGAGETPHMLLRPHVFMPEVTPDMYLLPRVNOGIIINPMLS 1140
QY 1098 DETNICNGKVPDGLTTLRNGTLVAFRGHVFWMLSPEPSPAPRIITEVWGISPIDTFTV 1157
Db 1141 DETNICNGKVPDGLTTLRNGTLVAFRGHVFWMLSPEPSPAPRIITEVWGISPIDTFTV 1200
QY 1158 RNCCEKTTFFKDSQVWRFTNDIKDAGYPKPIFGFGLTGQIVAAALSTAKYNWPESVY 1217
Db 1201 RNCCEKTTFFKDSQVWRFTNDIKDAGYPKPIFGFGLTGQIVAAALSTAKYNWPESVY 1260
QY 1218 FFKRGSSIOQYIYKQBPVKCGRRPALNYPVYGMTQVRRRFRFAIGPSQTHIRIQY 1277
Db 1261 FFKRGSSIOQYIYKQBPVKCGRRPALNYPVYGMTQVRRRFRFAIGPSQTHIRIQY 1320
QY 1278 SPARLAYQDKVLHNEVKVSIILWRGILPNVVTSAISLPNIRKPDGDYDYAFSKDQYNNIDV 1337
Db 1321 SPARLAYQDKVLHNEVKVSIILWRGILPNVVTSAISLPNIRKPDGDYDYAFSKDQYNNIDV 1380

QY 1338 PSRTARAITTRSGQTLSKWYNCP 1361
Db 1381 PSRTARAITTRSGQTLSKWYNCP 1404
RESULT 5
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124.557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Iuann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-124-557-62

Query Match 99.8%; Score 7308.7; DB 13; Length 1404;
Best Local Similarity 96.9%; Pred. No. 3.9e-169;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KRVCCTAELSCRCRCESFERGECDCDQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
Db 61 KRVCCTAELSCRCRCESFERGECDCDQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 121 PPSGASQITKSTTKRSPKPNKKTKVIESFEITE----- 156
Db 121 PPSGASQITKSTTKRSPKPNKKTKVIESFEITE----- 156

| | | | |
|----|------|---|------|
| QY | 157 | -----VKONKQNRKTKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST | 197 |
| Db | 181 | KIKSSKNSAANRELQKLKVKONKQNRKTKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST | 240 |
| QY | 198 | TOHNKYSTPKITAKPIINRPSLPNSDTSKETSITVNKETTVEKETTINKQISTDG | 257 |
| Db | 241 | TOHNKYSTPKITAKPIINRPSLPNSDTSKETSITVNKETTVEKETTINKQISTDG | 300 |
| QY | 258 | KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTPKPEPAS | 317 |
| Db | 301 | KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTPKPEPAS | 360 |
| QY | 318 | TTPEKPTPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBP | 377 |
| Db | 361 | TTPEKPTPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBP | 420 |
| QY | 378 | APTTPKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTK | 437 |
| Db | 421 | APTTPKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTK | 480 |
| QY | 438 | EPAPTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT | 497 |
| Db | 481 | EPAPTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPT | 540 |
| QY | 498 | TTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKKP | 557 |
| Db | 541 | TTKSAPTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKKP | 600 |
| QY | 558 | APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTT | 617 |
| Db | 601 | APTAPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTT | 660 |
| QY | 618 | PEEPAPTTPKAAAPNTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP | 677 |
| Db | 661 | PEEPAPTTPKAAAPNTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBP | 720 |
| QY | 678 | APTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTKG | 737 |
| Db | 721 | APTTPKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTKG | 780 |
| QY | 738 | TAPTTLKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTPK | 797 |
| Db | 781 | TAPTTLKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTPK | 840 |
| QY | 798 | KPAPTTPETPTPTSEVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTK | 857 |
| Db | 841 | KPAPTTPETPTPTSEVSTPTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTTKBPAPTTK | 900 |
| QY | 858 | TKTPAATKPEMTTAKDKTTERDARTPETTTAAPKMTKETATTTKTESKITATTTQV | 917 |
| Db | 901 | TKTPAATKPEMTTAKDKTTERDARTPETTTAAPKMTKETATTTKTESKITATTTQV | 960 |
| QY | 918 | TSTTTQDTPPKITTLTKTTTLAPKVTTTKITTEIMNKPEETAKPKDRATNSKATTPK | 977 |
| Db | 961 | TSTTTQDTPPKITTLTKTTTLAPKVTTTKITTEIMNKPEETAKPKDRATNSKATTPK | 1020 |
| QY | 978 | POKPTKAPKPTSTKPKTPRVRKPTTPTRKMTSTWPELNPSTRIAEAMLOTTTRPN | 1037 |
| Db | 1021 | POKPTKAPKPTSTKPKTPRVRKPTTPTRKMTSTWPELNPSTRIAEAMLOTTTRPN | 1080 |
| QY | 1038 | QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMXYLPRVNOGIIINPMLS | 1097 |
| Db | 1081 | QTPNSKLVENPKSEDAGAGETPHMLLRPHVFMPEVTPDMXYLPRVNOGIIINPMLS | 1140 |
| QY | 1098 | DETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSIDTFT | 1157 |
| Db | 1141 | DETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITEVWGIPSIDTFT | 1200 |
| QY | 1158 | RCNCEGKTFPFDKDSQVWRFTNDIKDAGYPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY | 1217 |
| Db | 1201 | RCNCEGKTFPFDKDSQVWRFTNDIKDAGYPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY | 1260 |

RESULT 6

US-10-124-557-48

Sequence 48, Application US/10124557

Publication No. US20020137894A1

GENERAL INFORMATION:

APPLICANT: Turner, Katherine C.

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 1354 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 48:

US-10-124-557-48

Query Match 96.7%; Score 7084.3; DB 13; Length 1354;

Best Local Similarity 97.0%; Pred. No. 1e-163;

Matches 1320; Conservative 8; Mismatches 26; Indels 7; Gaps 1;

QY 1 MAWKTLPIYLLILLISVFVIQVSSQDLSSCAGRCGEGYSRATCNCDYNCQHYMECCPDF 60

DB 1 MAWKTLPIYLLILLISVFVIQVSSQDLSSCAGRCGEGYSRATCNCDYNCQHYMECCPDF 60

```
QY 61 KRVTAEISKGRGCFESPERGECDCDAQCCKYDKCCDYBSCAEVHNPTSPSSKKAP 120
Db 61 KRVTAEISKGRGCFESPERGECDCDAQCCKYDKCCDYBSCAEVHNPTSPSSSSSS 120
QY 121 PPSGASQIKSTTKRSPKPNKKTKKVIESEEITEVKNKNKRTKKKPTPKPPVVDAG 180
Db 121 SSSSSSTIWKIKSGKNSKAANRELQKKL-----KVKDNKNKRTKKKPTPKPPVVDAG 173
QY 181 SGLDNGDFKVTPTDSTSTOHNKVSTSPKITTAKPINRPSLPNPSDSKETSLTVNKETT 240
Db 174 SGLDNGDFKVTPTDSTSTOHNKVSTSPKITTAKPINRPSLPNPSDSKETSLTVNKETT 233
QY 241 VETKETTITNKOTSDGKEKTTSAKETOSIEKTSIAKOLAPTSKVLAKPTKPAETTTKGA 300
Db 234 VETKETTITNKOTSDGKEKTTSAKETOSIEKTSIAKOLAPTSKVLAKPTKPAETTTKGA 293
QY 301 LTTPKETPTTPKEPASSTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKAPATTKE 360
Db 294 LTTPKETPTTPKEPASSTPKETPTTIKSAPTTPKEPAPTTTKSAPTTPKAPATTKE 353
QY 361 PAPTTPKEPAPTTTKAPATTTKSAPTTPKAPATTTPKEPAPTTPKAPATTKEPTTT 420
Db 354 PAPTTPKEPAPTTTKAPATTTKSAPTTPKAPATTTPKEPAPTTPKAPATTKEPTTT 413
QY 421 PKEPAPTTKEPAPTTPKAPATTAPKAPATTTPKEPAPTTPKAPATTTKPSPTTPKEPA 480
Db 414 PKEPAPTTKEPAPTTPKAPATTAPKAPATTTPKEPAPTTPKAPATTTKPSPTTPKEPA 473
QY 481 PTTTKSAPTTTKAPATTTKSAPTTPKPSPTTKKEPAPTTPKAPATTTPKEPAPTTKE 540
Db 474 PTTTKSAPTTTKAPATTTKSAPTTPKPSPTTKKEPAPTTPKAPATTTPKEPAPTTKE 533
QY 541 PAPTTPKEPAPTTTKAPATTAPKAPATTTPKEPAPTTPKKLPTTPKEKAPATTPEKAPT 600
Db 534 PAPTTPKEPAPTTTKAPATTAPKAPATTTPKEPAPTTPKKLPTTPKEKAPATTPEKAPT 593
QY 601 TPBELAPTTPEPTTPPEPAPTTPKAAANPTPKAPATTTPKEPAPTTPKAPATTKE 660
Db 594 TPBELAPTTPEPTTPPEPAPTTPKAAANPTPKAPATTTPKEPAPTTPKAPATTKE 653
QY 661 TAPTTPKGATPTLKAPATTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGATPTPK 720
Db 654 TAPTTPKGATPTLKAPATTPKKAPKELAPTTTKEPTSTTSKAPATTTPKGATPTPK 713
QY 721 EPAPTTPKEPAPTTPKGATPTLKAPATTPKKAPKELAPTTTKEPTSTTSKAPATTTP 780
Db 714 EPAPTTPKEPAPTTPKGATPTLKAPATTPKKAPKELAPTTTKEPTSTTSKAPATTTP 773
QY 781 KETAPTTKEPAPTTPKKAPATTPEPTTPPTTSVSTPTTTKEPTTIHKSDESTPELSAE 840
Db 774 KETAPTTKEPAPTTPKKAPATTPEPTTPPTTSVSTPTTTKEPTTIHKSDESTPELSAE 833
QY 841 PTPKALENSPKERGVTPTKPAATKPSMTTAKDKTERDLRTPETTTAAPKWTKETAT 900
Db 834 PTPKALENSPKERGVTPTKPAATKPSMTTAKDKTERDLRTPETTTAAPKWTKETAT 893
QY 901 TTEKTTESKITATTTQVSTSTTODTTPFKITLLKTTTLAPKVTTTKKTIITTIMKNKPEE 960
Db 894 TTEKTTESKITATTTQVSTSTTODTTPFKITLLKTTTLAPKVTTTKKTIITTIMKNKPEE 953
QY 961 TAKPKDRATNSKATTPKPKQKPTKAPKKPTSTSKPKTMRVRKPKTTPTPKMTSTMPELN 1020
Db 954 TAKPKDRATNSKATTPKPKQKPTKAPKKPTSTSKPKTMRVRKPKTTPTPKMTSTMPELN 1013
QY 1021 PPSRIAEAMLQTTTRNQTPNSKLVENVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMD 1080
Db 1014 PPSRIAEAMLQTTTRNQTPNSKLVENVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMD 1073
QY 1081 YLPRVFNQGIILINPMLSDETNICNGKPVQGLTTLRNGTLVAFRGHYFWMLSPSPSPAR 1140
Db 1074 YLPRVFNQGIILINPMLSDETNICNGKPVQGLTTLRNGTLVAFRGHYFWMLSPSPSPAR 1133
```

```
QY 1141 RITVWVGIPSPIDVTFRNCCEGTFPKDSQYWRFTNDIKDAGYPKDIFKGFGLTGOI 1200
Db 1134 RITVWVGIPSPIDVTFRNCCEGTFPKDSQYWRFTNDIKDAGYPKDIFKGFGLTGOI 1193
QY 1201 VAALSTAKYKNWPSVVFYFKRGGSIQOYIYKQEPVQKCPGRRPALNYPVYGMTCVRRRR 1260
Db 1194 VAALSTAKYKNWPSVVFYFKRGGSIQOYIYKQEPVQKCPGRRPALNYPVYGMTCVRRRR 1253
QY 1261 PERAIGSQTHIRIQSPARLAYQDKGVHLNEVKVLSILMRGLPNVVTSAISLNRKPD 1320
Db 1254 PERAIGSQTHIRIQSPARLAYQDKGVHLNEVKVLSILMRGLPNVVTSAISLNRKPD 1313
QY 1321 GYDYVAESKOYNNIDVPSRTARAITTRSGOTLSKQWYNCP 1361
Db 1314 GYDYVAESKOYNNIDVPSRTARAITTRSGOTLSKQWYNCP 1354
```

RESULT 7

```
US-10-124-557-46
; Sequence 46, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; City: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46
```

Query Match 96.4%; Score 7058.9; DB 13; Length 1320;
Best Local Similarity 97.0%; Pred. No. 4,1e-163;
Matches 1320; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

Query Match 96.4%; Score 7058.9; DB 13; Length 1320;
Best Local Similarity 97.0%; Pred. No. 4, 1e-163;
Matches 1320; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
QY 1 MAWKTLPIYLLLSLVFVIQVSSODLSLSCAGGEGYSDATCNCYHMECCPDF 60
Db 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
QY 61 KRVCATLSCGRCPESPERGECDDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120
Db 26 -----ELSCGRCPESPERGECDDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
QY 121 PPSGASQTIKSTTKRSPKPNKKTTKVIESEIEIEVDKNKRNKTKKPTPKPPVWDEAG 180
Db 80 PPSGASQTIKSTTKRSPKPNKKTTKVIESEIEIEVDKNKRNKTKKPTPKPPVWDEAG 139
QY 181 SGLDNGDFKVTPTDSTTQHNKVTSPKITTAKPINPRPSLPNPSDTSKETSITVKNKETT 240
Db 140 SGLDNGDFKVTPTDSTTQHNKVTSPKITTAKPINPRPSLPNPSDTSKETSITVKNKETT 199
QY 241 VETKEITTNKQTSIDGKEKTTSAKETOSIEKTSKAKOLAPTSKVLAKPTPKAETTTKGA 300
Db 200 VETKEITTNKQTSIDGKEKTTSAKETOSIEKTSKAKOLAPTSKVLAKPTPKAETTTKGA 259
QY 301 LTTPEKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKGPAPTTPKSAPTTPKEPAPTTPKE 360
Db 260 LTTPEKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKGPAPTTPKSAPTTPKEPAPTTPKE 319
QY 361 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKGPAPTTPKSAPTTPKGPAPTTPKGPAPTTP 420
Db 320 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKEPTPT 379
QY 421 PKEPAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKE 480
Db 380 PKEPAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKE 439
QY 481 PTTTKSAPTTPKEPAPTTPKSAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKE 540
Db 440 PTTTKSAPTTPKEPAPTTPKSAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKE 499
QY 541 PAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 600
Db 500 PAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 559
QY 601 TPBELAPTPEEPTPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKE 660
Db 560 TPBELAPTPEEPTPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKE 619
QY 661 TAPTTPKGTAPTTLKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 720
Db 620 TAPTTPKGTAPTTLKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 679
QY 721 EPAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 780
Db 680 EPAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 739
QY 781 KETAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 840
Db 740 KETAPTTPKEPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT 799
QY 841 PTPKALENPKGEGVPTTKTTPAATKEMTTAKDKTTERDLRTTPETTTAAAPKMTKETAT 900
Db 800 PTPKALENPKGEGVPTTKTTPAATKEMTTAKDKTTERDLRTTPETTTAAAPKMTKETAT 859
QY 901 TTEKTTESKITATTQVSTTTQDTPFKITLTKTTLAPKVTITTKITITIEIMNKPEE 960
Db 860 TTEKTTESKITATTQVSTTTQDTPFKITLTKTTLAPKVTITTKITITIEIMNKPEE 919
QY 961 TAKPKDRATNSKATTPKPKOKPTKAPKPKSTKPKTTPRVRPKTTPTRKMTSTMPELN 1020
Db 920 TAKPKDRATNSKATTPKPKOKPTKAPKPKSTKPKTTPRVRPKTTPTRKMTSTMPELN 979

QY 1021 PTSRIAEMLOTTTRPNQTPNSKLVENPKSEADAGAGETPHMLLRPHVFMPEVTPDMD 1080
Db 980 PTSRIAEMLOTTTRPNQTPNSKLVENPKSEADAGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1081 YLPRVNPNGIILNPMLSDETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1140
Db 1040 YLPRVNPNGIILNPMLSDETNI CNKGPVDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1099
QY 1141 RITEVWGPISPIDVTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKEIFKGFGLTGOI 1200
Db 1100 RITEVWGPISPIDVTRCNCEGKTFPFKDSQYWRFTNDIKDAGYPKEIFKGFGLTGOI 1159
QY 1201 VAALSTAKYKNWPSVYFFKGGSIQYIYKQBPVKQKCPGRRPALNYPVYGEMTQVRRRR 1260
Db 1160 VAALSTAKYKNWPSVYFFKGGSIQYIYKQBPVKQKCPGRRPALNYPVYGEMTQVRRRR 1219
QY 1261 FERAIGPSQTHTRI QXSPARLAYQDKGLVHNEVKVSLILWRGLNPNVTSALSLPNIRKPD 1320
Db 1220 FERAIGPSQTHTRI QXSPARLAYQDKGLVHNEVKVSLILWRGLNPNVTSALSLPNIRKPD 1279
QY 1321 GYDYAFSKQOYXNIDVPSTRARAITTRSGQTLSSKVVWYINCP 1361
Db 1280 GYDYAFSKQOYXNIDVPSTRARAITTRSGQTLSSKVVWYINCP 1320

RESULT 9

US-10-124-557-42
; Sequence 42, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match          96.3%; Score 7049; DB 13; Length 1311;
Best Local Similarity 96.3%; Pred. No. 7e-163; Indels 50; Gaps 1;
Matches 1311; Conservative 0; Mismatches 0;

QY 1  MAMKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHMECCPDF 60
DB 1  MAMKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHMECCPDF 60

QY 61  KRVTAAELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESCAEVHNFTSPSSKKAP 120
DB 61  KRVTAAELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESCAEVHNFTSPSSKKAP 120

QY 61  KRVTAAELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESCAEVHNFTSPSSKKAP 120
DB 61  KRVTAAELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESCAEVHNFTSPSSKKAP 120

QY 121  PPSGASQTIKSTTKRSPKPNKKTKVIESEITEVKDNKKRTKKKPTKPPVVDKAP 180
DB 121  PPSGASQTIKSTTKRSPKPNKKTKVIESEITEVKDNKKRTKKKPTKPPVVDKAP 180

QY 181  SGLDNGDFKVTTPDSTTOHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 240
DB 181  SGLDNGDFKVTTPDSTTOHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETT 240

QY 241  VETKETTNNKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 300
DB 241  VETKETTNNKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGA 300

QY 301  LTTPKKEPTTTKPEPASSTTKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 360
DB 301  LTTPKKEPTTTKPEPASSTTKPEPTPTTKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 360

QY 361  PAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEPTT 420
DB 361  PAPTTKPEAPTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEPTT 420

QY 421  PKEAPTTKPEAPTTTKPEAPTTAPKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEPA 480
DB 421  PKEAPTTKPEAPTTTKPEAPTTAPKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEPA 480

QY 481  PTTTKSAPTTTKPEAPTTTKSAPTTKPEPSPTTKPEAPTTKPEAPTTTKKPAPTTPKE 540
DB 481  PTTTKSAPTTTKPEAPTTTKSAPTTKPEPSPTTKPEAPTTKPEAPTTTKKPAPTTPKE 540

QY 541  PAPTTKPEAPTTTKKPAPTAPKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 600
DB 541  PAPTTKPEAPTTTKKPAPTAPKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 600

QY 601  TPEELAPTTPEPTTPEEPAPTTPKAAANTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKE 660
DB 601  TPEELAPTTPEPTTPEEPAPTTPKAAANTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKE 660

QY 661  TAPTTKGTATTLKPEAPTTKPKAPKELAPTTTKEPTSTTSOKPAPTTKGTAPTTTPK 720
DB 661  TAPTTKGTATTLKPEAPTTKPKAPKELAPTTTKEPTSTTSOKPAPTTKGTAPTTTPK 720

QY 721  EPAPTTKPEAPTTKGTAPTTLKEAPATTTPKKAPKELAPTTTKEPTSTTSOKPAPTT 780
DB 721  EPAPTTKPEAPTTKGTAPTTLKEAPATTTPKKAPKELAPTTTKEPTSTTSOKPAPTT 780

QY 781  KETAPTTKPEAPTTKPEAPTTKPEPTTSEVSTPTTKEPTTIHKSPESTPELSAE 840
DB 781  KETAPTTKPEAPTTKPEAPTTKPEPTTSEVSTPTTKEPTTIHKSPESTPELSAE 840

QY 841  PPKALENSPKGVPVTTTKTAAATKPEMTTTAKDKTTERDLRTTPTTTAAAPMTKETAT 900
DB 841  PPKALENSPKGVPVTTTKTAAATKPEMTTTAKDKTTERDLRTTPTTTAAAPMTKETAT 900

QY 901  TTEKTTESKIDATTTQVSTTTQDTTPPKIITLKTTLAPKVTTTKITTTTEIMNKPEE 960
DB 901  TTEKTTESKIDATTTQVSTTTQDTTPPKIITLKTTLAPKVTTTKITTTTEIMNKPEE 960

QY 961  TAKPKDRATNSKATTPKPKQKPTSTTKPKTAPRVRKPTTTPPKMTSTWPELN 1020
DB 961  TAKPKDRATNSKATTPKPKQKPTSTTKPKTAPRVRKPTTTPPKMTSTWPELN 1020

QY 971  PTSTRIAEAMLOTTTRPNQTPNSKLVNPKSESDAGABGETPHMLLRPHVMEVEVTDMD 1080
DB 971  PTSTRIAEAMLOTTTRPNQTPNSKLVNPKSESDAGABGETPHMLLRPHVMEVEVTDMD 1080

QY 1081  YLPRVNOGIIINPMLSDETINICNGKPVDLGTLTTLRNGTLVAFRGHYFWMLSPEFSPSPAR 1140
DB 1081  YLPRVNOGIIINPMLSDETINICNGKPVDLGTLTTLRNGTLVAFRGHYFWMLSPEFSPSPAR 1140

QY 1141  RITEVWGIISPIDTITRNCNCEKTEFFKDSQYWRFTNDIKDAGYKPKIFKGFGLTGQI 1200
DB 1141  RITEVWGIISPIDTITRNCNCEKTEFFKDSQYWRFTNDIKDAGYKPKIFKGFGLTGQI 1200

QY 1201  VAALSTAKYKNWPESVYFFKRGSGIOQYIKQBPVQKCPGRRPALNPNVYGEWTOVRRR 1260
DB 1201  VAALSTAKYKNWPESVYFFKRGSGIOQYIKQBPVQKCPGRRPALNPNVYGEWTOVRRR 1260

QY 1261  FERRAIGPSQTHIRIOYSPARLAYQDKGVHLHNEVKVLSILWRGLPNVVTSAISLPIRKP 1320
DB 1261  FERRAIGPSQTHIRIOYSPARLAYQDKGVHLHNEVKVLSILWRGLPNVVTSAISLPIRKP 1320

QY 1321  GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1361
DB 1321  GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1361

QY 1361  GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1361
DB 1361  GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1361

RESULT 10
US-10-124-557-52
; Sequence 52, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851

```

```
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match          96.2%; Score 7044.6; DB 13; Length 1363;
Best Local Similarity 94.0%; Pred. No. 9.4e-163;
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYRSDATCNCVNCQHYWECPPDF 60
Db 1 MAWKTLPIYLLLSLVFVIQVSSQ-----25

QY 61 KRVCATLSCGRGCFESFERGRCDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
Db -----ELSCGRCGCFESFERGRCDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79

QY 121 PPASGASOTIKSTTKRSKPKPNKKTKKVIIESEITE-----156
Db 80 PPASGASOTIKSTTKRSKPKPNKKTKKVIIESEITEEHSVSENQESSSSSSSSSTIW 139

QY 157 -----VKDNKNKRTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 197
Db 140 KIKSSKNSAANRELQKKLVKDNKNKRTKKKPTPKPPVWDEAGSLDNGDFKVTTPDTST 199

QY 198 TQHNKYSTSPKIITAKIPNRPSPSPNSDTSKETSITLVNKETTVEKETTINKQISTDG 257
Db 200 TQHNKYSTSPKIITAKIPNRPSPSPNSDTSKETSITLVNKETTVEKETTINKQISTDG 259

QY 258 KEKTTISAKETQSLEKTSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 317
Db 260 KEKTTISAKETQSLEKTSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 319

QY 318 TTPKEPTTPTIKSAPTTTKEPAPTPTTKSAPTTTKEPAPTPTTKKEPAPTPTTKKEP 377
Db 320 TTPKEPTTPTIKSAPTTTKEPAPTPTTKSAPTTTKEPAPTPTTKKEPAPTPTTKKEP 379

QY 378 APITTKSAPTTTKEPAPTPTTKPAPTTPKEPAPTTPKEPTTTPKEPAPTTPKEPAPTTPK 437
Db 380 APITTKSAPTTTKEPAPTPTTKPAPTTPKEPAPTTPKEPTTTPKEPAPTTPKEPAPTTPK 439

QY 438 EPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 497
Db 440 EPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 499

QY 498 TTKSAPTTTPKEPSPTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 557
Db 500 TTKSAPTTTPKEPSPTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 559

QY 558 APTAPKEPAPTTPKEAPTTPKLPITTPTEKLAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 617
Db 560 APTAPKEPAPTTPKEAPTTPKLPITTPTEKLAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 619

QY 618 PEEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKG 737
Db 620 PEEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKG 739

QY 678 APTTPKKAPKELAPTTPKEPTSTTSKDPAPTTPKGAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 797
Db 680 APTTPKKAPKELAPTTPKEPTSTTSKDPAPTTPKGAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 799

QY 738 TAPPTLKAPAPTTPKKAPKELAPTTPKGPTSTTSKDPAPTTPKETAPTTPKEPAPTTPK 797
Db 740 TAPPTLKAPAPTTPKKAPKELAPTTPKGPTSTTSKDPAPTTPKETAPTTPKEPAPTTPK 799

QY 798 KPAPTTPPTTSEVSTPTTKEPTTIHKSPTDESTPELSPEPTPKALENSPKPGVPT 857
Db 800 KPAPTTPPTTSEVSTPTTKEPTTIHKSPTDESTPELSPEPTPKALENSPKPGVPT 859
```

```
QY 858 TKTPAATKPEMTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTESKITATTQV 917
Db 860 TKTPAATKPEMTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTESKITATTQV 919

QY 918 TSITTTQDTTPPKIITLTKTTLAPKVTITTKTITTEIMNKPEETAKPKORATNSKATTPK 977
Db 920 TSITTTQDTTPPKIITLTKTTLAPKVTITTKTITTEIMNKPEETAKPKORATNSKATTPK 979

QY 978 PQKETKAPKPTSTKPKTMTPRVRKPKTTPTRKWTSTMPDELNPTSIAEAMLOTTTRPN 1037
Db 980 PQKETKAPKPTSTKPKTMTPRVRKPKTTPTRKWTSTMPDELNPTSIAEAMLOTTTRPN 1039

QY 1038 QTNSKLIVNPKSEADAGGAGETPHMLLRPHVMPVETPDMDYLPVPNQGIINPMLS 1097
Db 1040 QTNSKLIVNPKSEADAGGAGETPHMLLRPHVMPVETPDMDYLPVPNQGIINPMLS 1099

QY 1098 DETNICNGKPGVDGLTTLNRNGTLVAFRGHYFWMILSPSPSPARRITEVWGISPIDTVFT 1157
Db 1100 DETNICNGKPGVDGLTTLNRNGTLVAFRGHYFWMILSPSPSPARRITEVWGISPIDTVFT 1159

QY 1158 RCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIWAALSTAKYKNWPESVY 1217
Db 1160 RCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIWAALSTAKYKNWPESVY 1219

QY 1218 FFKEGSGIOQYIYKQEPVQKCPGRRPALNYPVYGEVTQVRRRRFERAIGPSQTHIRIQY 1277
Db 1220 FFKEGSGIOQYIYKQEPVQKCPGRRPALNYPVYGEVTQVRRRRFERAIGPSQTHIRIQY 1279

QY 1278 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAASKQYQYINIDV 1337
Db 1280 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAASKQYQYINIDV 1339

QY 1338 PSTARAITTRSGQTLTKVWYNCP 1361
Db 1340 PSTARAITTRSGQTLTKVWYNCP 1363

RESULT 11
US-10-124-557-50
; Sequence 50, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
```

FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,922
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

Query Match 93.4%; Score 6836.3; DB 13; Length 1314;
Best Local Similarity 94.1%; Pred. No. 1e-157;
Matches 1281; Conservative 8; Mismatches 25; Indels 47; Gaps 2;

QY 1 MAWKTLPIYLLILLVSVFVQQVSSQDLSSCAGRCGEGYSRDATCNCQYHMECCPDF 60
DB 1 MAWKTLPIYLLILLVSVFVQQVSSQDLSSCAGRCGEGYSRDATCNCQYHMECCPDF 60

QY 61 KRYCTAELSCGRCFESFERGRCDDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 61 KRYCTAE-----HSYSENQESSSS 80

QY 121 PPSGASQTIKSTTKRSKPPNKKTKVIBSEBITVEKONKNRTKKKPTKPPVDEAG 180
DB 81 SSSSSSTTWIKSSXNSAANRELQKLI-----KVKONKNRTKKKPTKPPVDEAG 133

QY 181 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSVNKETT 240
DB 134 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSVNKETT 193

QY 241 VETKETTNNKQSTDKKETTSAKETQSTKTSKADLAPTSKVLAKPTPKAETTKGPA 300
DB 194 VETKETTNNKQSTDKKETTSAKETQSTKTSKADLAPTSKVLAKPTPKAETTKGPA 253

QY 301 LTTPEKPTTPKPEASSTTKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 360
DB 254 LTTPEKPTTPKPEASSTTKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 313

QY 361 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKE 420
DB 314 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKE 373

QY 421 PKEAPTTKBPAPTTTPKPEAPTTAPKAPPTPKPEAPTTTPKPEAPTTTKPEAPTTTKE 480
DB 374 PKEAPTTKBPAPTTTPKPEAPTTAPKAPPTPKPEAPTTTPKPEAPTTTKPEAPTTTKE 433

QY 481 PTTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKE 540
DB 434 PTTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKE 493

QY 541 PAPTTPKPEAPTTTKPEAPTTAPKAPPTPKPEAPTTTPKPEAPTTTKPEAPTTTKE 600
DB 494 PAPTTPKPEAPTTTKPEAPTTAPKAPPTPKPEAPTTTPKPEAPTTTKPEAPTTTKE 553

QY 601 TPELAPTTPEPTPTTPPEAPTTPKAAANTPKPEAPTTTPKPEAPTTTKPEAPTTTKE 660
DB 554 TPELAPTTPEPTPTTPPEAPTTPKAAANTPKPEAPTTTPKPEAPTTTKPEAPTTTKE 613

QY 661 TAPTTPKGTATTLKPEAPTTPKKAPKELAPTTTKPEPTSTTSOKPAPTTPKGTAPTTPK 720
DB 614 TAPTTPKGTATTLKPEAPTTPKKAPKELAPTTTKPEPTSTTSOKPAPTTPKGTAPTTPK 673

QY 721 EPAPTTTPKPEAPTTTPKGTAPTTLKEPAPTTTPKPKAPKELAPTTTKGPTSTTSOKPAPTT 780
DB 674 EPAPTTTPKPEAPTTTPKGTAPTTLKEPAPTTTPKPKAPKELAPTTTKGPTSTTSOKPAPTT 733

QY 781 KETAPTTPKEPAPTTPKKPAPTTPPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAE 840
DB 734 KETAPTTPKEPAPTTPKKPAPTTPPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAE 793

QY 841 PTPKALENSPKPEPGVPTTKTAPATKPEMTTTAKDKTTERDLRTTETTTTAAAPMKETAT 900
DB 794 PTPKALENSPKPEPGVPTTKTAPATKPEMTTTAKDKTTERDLRTTETTTTAAAPMKETAT 853

QY 901 TTEKTTESKLTATTTTQVSTTTTQDTPPKITTLTKTTLAPKVTTKTKTTTTEIMNKPEE 960
DB 854 TTEKTTESKLTATTTTQVSTTTTQDTPPKITTLTKTTLAPKVTTKTKTTTTEIMNKPEE 913

QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMTPEVRKPKTPTTPRKMSTMPELN 1020
DB 914 TAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMTPEVRKPKTPTTPRKMSTMPELN 973

QY 1021 PTSRIAAMLOTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFPVEVTPDMD 1080
DB 974 PTSRIAAMLOTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFPVEVTPDMD 1033

QY 1081 YLPRVNPQGIILNPMLSDETNIENCGKPVVDGLTTLRNGTLVAPRGHYFWMLSPEFPPSPAR 1140
DB 1034 YLPRVNPQGIILNPMLSDETNIENCGKPVVDGLTTLRNGTLVAPRGHYFWMLSPEFPPSPAR 1093

QY 1141 RITEVWGIPIPIIDTVTRCNCEGKTPFFKDSQVWRFNDIKDAGYKPIFKGFGGUTGQI 1200
DB 1094 RITEVWGIPIPIIDTVTRCNCEGKTPFFKDSQVWRFNDIKDAGYKPIFKGFGGUTGQI 1153

QY 1201 VAALSTAKYKNPESVYFFKRGSGIOQYIKQEPVQKCPGRRPALNYPVYVGMTQVRRRR 1260
DB 1154 VAALSTAKYKNPESVYFFKRGSGIOQYIKQEPVQKCPGRRPALNYPVYVGMTQVRRRR 1213

QY 1261 FERAIGPSQTHITRIQVSPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKP 1320
DB 1214 FERAIGPSQTHITRIQVSPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKP 1273

QY 1321 GYDYAFSDQYNNIDVPSRTARAITTRSGOTLSKVWYNCP 1361
DB 1274 GYDYAFSDQYNNIDVPSRTARAITTRSGOTLSKVWYNCP 1314

RESULT 12
US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114

```

; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserit, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Query Match          93.1%; Score 6820.2; DB 13; Length 1313;
Best Local Similarity 94.0%; Pred. No. 2.5e-157;
Matches 1279; Conservative 8; Mismatches 26; Indels 48; Gaps 2;

QY 1 MAWKTLPIYLLLSLVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIQQVSSQ-----25
QY 61 KRVCIAELSCGKCFEFREGRECDCAQCKYDKCCPDYSEFCAEVNHTSPSSKAP 120
DB 26 -----ELSCGKCFEFREGRECDCAQCKYDKCCPDYSEFCAEHSVSENQSSSS 79
QY 121 PPSGASQIKSTTKRSPKPNKKTKKXVIESEITEVKDNKKRTKKKTPKPPVDEAG 180
DB 80 SSSSSSTLIWKSSKNSAANRELQKL-----KVDKNKKRTKKKTPKPPVDEAG 132
QY 181 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINPRPSLPNSDTSKETSITVNKETT 240
DB 133 SGLDNGDFKVTTPDSTTQHNKYSTSPKITTAKPINPRPSLPNSDTSKETSITVNKETT 192
QY 241 VETKEITTNKQSTDGKKEITSAKETQSIEKTSADLAPTSKVLAKPPTKAPTITKGP 300
DB 193 VETKEITTNKQSTDGKKEITSAKETQSIEKTSADLAPTSKVLAKPPTKAPTITKGP 252
QY 301 LTPPKETPTTPKPEASTTPKPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 360
DB 253 LTPPKETPTTPKPEASTTPKPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 312
QY 361 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTPKKPAPTTPKPEAPTTKPTPT 420
DB 313 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTPKKPAPTTPKPEAPTTKPTPT 372
QY 421 PKPEAPTTKPEAPTTKPEAPTTAPKPAPTTPKPEAPTTKPEAPTTKPEAPTTTKE 480
DB 373 PKPEAPTTKPEAPTTKPEAPTTAPKPAPTTPKPEAPTTTKEAPTTTKEAPTTTKE 432
QY 481 PTTTTSAPTTTKEAPTTTTSAPTTTKEPSPTTTKEAPTTTKEAPTTTKEAPTTTKE 540
DB 433 PTTTTSAPTTTKEAPTTTTSAPTTTKEPSPTTTKEAPTTTKEAPTTTKEAPTTTKE 492
QY 541 PAPTTPKPEAPTTTKKPAPTAPKPAPTTPKETAPTTPKKLAPTTPKELAPTTPKPA 600
DB 493 PAPTTPKPEAPTTTKKPAPTAPKPAPTTPKETAPTTPKKLAPTTPKELAPTTPKPA 552
QY 601 TPEELAPTTPPEPTPTTPPEAPTTPKAAAPNTPKPEAPTTKPEAPTTKPEAPTTTKE 660
DB 553 TPEELAPTTPPEPTPTTPPEAPTTPKAAAPNTPKPEAPTTTKEAPTTTKEAPTTTKE 612
QY 661 TAPTTPKGTAPTTLKPEAPTTTPKKAPKELAPTTTKEPTSTTSKAPPTPKGTAPTTPK 720
DB 613 TAPTTPKGTAPTTLKPEAPTTTPKKAPKELAPTTTKEPTSTTSKAPPTPKGTAPTTPK 672

```

```

QY 721 EPAPTTPKPEAPTTTPKGTAPTTLKPEAPTTTPKKAPKELAPTTTTPKPTSTTSKAPPTTP 780
DB 673 EPAPTTPKPEAPTTTPKGTAPTTLKPEAPTTTPKKAPKELAPTTTTPKPTSTTSKAPPTTP 732
QY 781 KETAPTTPKPEAPTTTPKAPPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTELSAAE 840
DB 733 KETAPTTPKPEAPTTTPKAPPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTELSAAE 792
QY 841 PTPKALENSPKPEPGVPTTTPKTAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKETAT 900
DB 793 PTPKALENSPKPEPGVPTTTPKTAATKPEMTTAKDKTTERDLRTTTPETTTAAAPKMTKETAT 852
QY 901 TTEKTTESKITATTTOVTSTTTQDTTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEE 960
DB 853 TTEKTTESKITATTTOVTSTTTQDTTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEE 912
QY 961 TAKPKDRATNSKATTPKPKPTKAPKPTSPKPKTTPKPKPTTPKPKPTTPKPKPTTPKPKPT 1020
DB 913 TAKPKDRATNSKATTPKPKPTKAPKPTSPKPKTTPKPKPTTPKPKPTTPKPKPTTPKPKPT 972
QY 1021 PTSRIABAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPDMD 1080
DB 973 PTSRIABAMLQTTTRPNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHVFMPEVTPDMD 1032
QY 1081 YLPRVNPQGIILINPMLSDETNI CNKPKVDGLTTLRNGLTVAFRGHYFWMLSPPSPSPAR 1140
DB 1033 YLPRVNPQGIILINPMLSDETNI CNKPKVDGLTTLRNGLTVAFRGHYFWMLSPPSPSPAR 1092
QY 1141 RITEVWGIPIPIDTFTVTRCNCCKTFFPKDQSYWRFTNDIKDAGYKPKIPIKFGFGLTGQI 1200
DB 1093 RITEVWGIPIPIDTFTVTRCNCCKTFFPKDQSYWRFTNDIKDAGYKPKIPIKFGFGLTGQI 1152
QY 1201 VAALSTAKYKNWPESVYFFPKRGGSIQQYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRR 1260
DB 1153 VAALSTAKYKNWPESVYFFPKRGGSIQQYIYKQBPVQKCPGRRPALNYPVVGEMTQVRRR 1212
QY 1261 FERAIGPSQHTTIRIQYSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPD 1320
DB 1213 FERAIGPSQHTTIRIQYSPARLAYQDKVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPD 1272
QY 1321 GYDYAFSKDQYINIDVPSTARAITTRSGOTLSKVWYNCP 1361
DB 1273 GYDYAFSKDQYINIDVPSTARAITTRSGOTLSKVWYNCP 1313

```

RESULT 13

US-10-124-557-44

; Sequence 44, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Geener, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/124,557

; FILING DATE: 16-Apr-2002

; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/643,502
;; FILING DATE: 18-JAN-1991
;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cserr, Luann
;; REGISTRATION NUMBER: 31,822
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 876-1170
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 44:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1270 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

Query Match 92.7%; Score 6784.9; DB 13; Length 1270;
Best Local Similarity 93.3%; Pred. No. 1.7e-156;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLSVFIQVVSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFIQVVSQ----- 25
QY 61 KRVTAEISCKGRCFESFERECDDAQCKYDKCCPDYESFOAEVHNFTSPSSKAP 120
DB 26 -----ELSCGRCFESFERECDDAQCKYDKCCPDYESFCA----- 64
QY 121 PPGASQTIKSTTKRSPKPNKXKTKVIESEETEVKDNKNRTKKKPTKPPVVDK 180
DB 65 -----EVKDNKNRTKKKPTKPPVVDK 89
QY 181 SGLDNGDKVTPDTSTQHNKVSSTPKITTAKEINPRPSLPNSDTSKETSLSVWKETT 240
DB 90 SGLDNGDKVTPDTSTQHNKVSSTPKITTAKEINPRPSLPNSDTSKETSLSVWKETT 149
QY 241 VETKETITTNKQTSIDGKEKITSKETSIEKTSKAKOLAPTSKVLAKPTPKAETTKGPA 300
DB 150 VETKETITTNKQTSIDGKEKITSKETSIEKTSKAKOLAPTSKVLAKPTPKAETTKGPA 209
QY 301 LTTPEKPTTTPKGPASTTPKEPTTTIKSAPTTPEKAPTTTKSAPTTTKSAPTTTKE 360
DB 210 LTTPEKPTTTPKGPASTTPKEPTTTIKSAPTTPEKAPTTTKSAPTTTKSAPTTTKE 269
QY 361 PAPTPKEPAPTTTKEPAPTTTKSAPTTPKPAPTTTPKAPTTTPKEPAPTTTKEPTTT 420
DB 270 PAPTPKEPAPTTTKEPAPTTTKSAPTTPKPAPTTTPKAPTTTPKEPAPTTTKEPTTT 329
QY 421 PKEPAPTTTKEPAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTKEPAPTTTKEP 480
DB 330 PKEPAPTTTKEPAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTKEPAPTTTKEP 389
QY 481 PTTTKSAPTTTKEPAPTTTKSAPTTPKPSPPTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540
DB 390 PTTTKSAPTTTKEPAPTTTKSAPTTPKPSPPTTKEPAPTTTKEPAPTTTKEPAPTTTKE 449
QY 541 PAPTPKGPAPTTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPEKAPPT 600
DB 450 PAPTPKGPAPTTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPKAPTTTPEKAPPT 509
QY 601 TPEELAPTTPEPTTTPPEAPTTTPKAAAPNTPKEPAPTTTPKAPTTTPKAPTTTPE 660
DB 510 TPEELAPTTPEPTTTPPEAPTTTPKAAAPNTPKEPAPTTTPKAPTTTPKAPTTTPE 569

QY 661 TAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPK 720
DB 570 TAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPK 629
QY 721 EPAPTTPEPAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTTP 780
DB 630 EPAPTTPEPAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTTP 689
QY 781 KETAPTTPEPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTTKEPTSTTSKDPAPTTTP 840
DB 690 KETAPTTPEPAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTTKEPTSTTSKDPAPTTTP 749
QY 841 PTPKALENSPKBPGVPTTKTTPAATKPEMTTAKDKTTERDLRTPPTTAAAPKMTKETAT 900
DB 750 PTPKALENSPKBPGVPTTKTTPAATKPEMTTAKDKTTERDLRTPPTTAAAPKMTKETAT 809
QY 901 TTEKTTESKITATTQVTTSTTQDTPFKITLTKTTTLPKVTTKKTIITTEIMNKPEE 960
DB 810 TTEKTTESKITATTQVTTSTTQDTPFKITLTKTTTLPKVTTKKTIITTEIMNKPEE 869
QY 961 TAKPKDRATNSKATTPKPOKPTKAPKPKPTSTKKPKTMPRVKPKPTTPPKMTSTMPELN 1020
DB 870 TAKPKDRATNSKATTPKPOKPTKAPKPKPTSTKKPKTMPRVKPKPTTPPKMTSTMPELN 929
QY 1021 PTSRIAEMLOTTTRPNQTPNSKLVEVNPKSEADAGAEGETPHMLLRPHVFMPEVTPDMD 1080
DB 930 PTSRIAEMLOTTTRPNQTPNSKLVEVNPKSEADAGAEGETPHMLLRPHVFMPEVTPDMD 989
QY 1081 YLPRVNPQGIINPMLSDETINICNGRPVDGLTTLRNGTLVAFRGHVFWMLSPPSPSPAR 1140
DB 990 YLPRVNPQGIINPMLSDETINICNGRPVDGLTTLRNGTLVAFRGHVFWMLSPPSPSPAR 1049
QY 1141 RITEVWGIPIPIDTFTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI 1200
DB 1050 RITEVWGIPIPIDTFTRCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI 1109
QY 1201 VAALSTAKYKNWPESVYFFKRGSGSIQYIYKQBPVQKCPGRRPALNTPVVGEMTQVRRR 1260
DB 1110 VAALSTAKYKNWPESVYFFKRGSGSIQYIYKQBPVQKCPGRRPALNTPVVGEMTQVRRR 1169
QY 1261 FERAIGSQTHTIRIQYSPARLAYQDKGVLHNEVKVSLMRLGLEPNVVTSAISLNPRIKPD 1320
DB 1170 FERAIGSQTHTIRIQYSPARLAYQDKGVLHNEVKVSLMRLGLEPNVVTSAISLNPRIKPD 1229
QY 1321 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1361
DB 1230 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKVWYNCP 1270

RESULT 14
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

| | | | |
|---|------|---|------|
| QY | 558 | A P T A P K E B A P P T P K E T A P T P K K L T P T T P R E K L A P T T P E K P A P T T P E L A P T T P E E P T P T T | 617 |
| DB | 601 | A P T A P K E B A P P T P K E T A P T P K K L T P T T P R E K L A P T T P E K P A P T T P E L A P T T P E E P T P T T | 660 |
| QY | 618 | P E E P A P T T P K A A A N T P K E B A P T T P K E B A P T T P K E B A P T T P K E T A P T T P K G T A P T T L K E P | 677 |
| DB | 661 | P E E P A P T T P K A A A N T P K E B A P T T P K E B A P T T P K E B A P T T P K E T A P T T P K G T A P T T L K E P | 720 |
| QY | 678 | A P T T P K K P A K E L A P T T T K E P T S T T S D K P A P T T P K G T A P T T P K G T A P T T P K E B A P T T P K E B A P T T P K G | 737 |
| DB | 721 | A P T T P K K P A K E L A P T T T K E P T S T T S D K P A P T T P K G T A P T T P K E B A P T T P K E B A P T T P K G | 780 |
| QY | 738 | T A P T T L K E B A P T T P K P A K E L A P T T T K G T S T T S D K P A P T T P K E T A P T T P K E T A P T T P K E B A P T T P K | 797 |
| DB | 781 | T A P T T L K E B A P T T P K P A K E L A P T T T K G T S T T S D K P A P T T P K E T A P T T P K E T A P T T P K E B A P T T P K | 840 |
| QY | 798 | K P A P T T P E T P P P T T S E V S T T T T K E P T T I H K S P E S T P E L S A E P T P K A L E N S P K E G V P T | 857 |
| DB | 841 | K P A P T T P E T P P P T T S E V S T T T T K E P T T I H K S P E S T P E L S A E P T P K A L E N S P K E G V P T | 900 |
| QY | 858 | T K T P A A T K P E M T T A K O K T T E R D L R T T P E T T A A P K M T K E T A T T T E K T T E S K I T A T T T Q V | 917 |
| DB | 901 | T K T P A A T K P E M T T A K O K T T E R D L R T T P E T T A A P K M T K E T A T T T E K T T E S K I T A T T T Q V | 960 |
| QY | 918 | T S T T T Q U T T F K I T T L K T T T L A P K V T T T K K I T T T E I M N K P B E T A K P D R A T N S K A T T P K | 977 |
| DB | 961 | T S T T T Q U T T F K I T T L K T T T L A P K V T T T K K I T T T E I M N K P B E T A K P D R A T N S K A T T P K | 1020 |
| QY | 978 | P Q K P T A P K K P T S T K K P K T M P R V R K P T T P P R K M T S T M P E L N P T S R I A E A M L Q T T T R E N | 1037 |
| DB | 1021 | P Q K P T A P K K P T S T K K P K T M P R V R K P T T P P R K M T S T M P E L N P T S R I A E A M L Q T T T R E N | 1080 |
| QY | 1038 | Q T P N S K L V E N P K S E D A G A E G E T P H M L L R P H V E M P E V T P D M D Y L P R V P N Q G I I I N P M L S | 1097 |
| DB | 1081 | Q T P N S K L V E N P K S E D A G A E G E T P H M L L R P H V E M P E V T P D M D Y L P R V P N Q G I I I N P M L S | 1140 |
| RESULT 15 | | | |
| US-10-124--557-74 | | | |
| ; Sequence 74, Application US/10124557 | | | |
| ; Publication No. US20020137894A1 | | | |
| ; GENERAL INFORMATION: | | | |
| APPLICANT: Turner, Katherine | | | |
| Clark, Stephen C. | | | |
| Jacobs, Kenneth | | | |
| Hewick, Rodney M. | | | |
| Gesner, Thomas G. | | | |
| TITLE OF INVENTION: Megakaryocyte Stimulating Factors | | | |
| NUMBER OF SEQUENCES: 143 | | | |
| CORRESPONDENCE ADDRESS: | | | |
| ADDRESSEE: Genetics Institute, Inc. | | | |
| STREET: 87 CambridgePark Drive | | | |
| CITY: Cambridge | | | |
| STATE: Massachusetts | | | |
| COUNTRY: U.S.A. | | | |
| ZIP: 02140 | | | |
| COMPUTER READABLE FORM: | | | |
| MEDIUM TYPE: Floppy disk | | | |
| COMPUTER: IBM PC compatible | | | |
| OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| SOFTWARE: PatentIn Release #1.0, Version #1.25 | | | |
| CURRENT APPLICATION DATA: | | | |
| APPLICATION NUMBER: US/10/124,557 | | | |
| FILING DATE: 16-Apr-2002 | | | |
| CLASSIFICATION: <Unknown> | | | |
| PRIOR APPLICATION DATA: | | | |
| APPLICATION NUMBER: US 07/643,502 | | | |
| FILING DATE: 18-JAN-1991 | | | |
| APPLICATION NUMBER: US 07/546,114 | | | |
| FILING DATE: 29-JUN-1990 | | | |
| APPLICATION NUMBER: US 07/457,196 | | | |
| FILING DATE: 29-DEC-1989 | | | |
| APPLICATION NUMBER: US 07/190,901 | | | |

| | | | | | | | | | | | |
|---|--|--|----|--|--|------|--|--|------|--|--|
| FILING DATE: 08-AUG-1989 | | | QY | | | KETA | | | 840 | | |
| ATTORNEY/AGENT INFORMATION: | | | Db | | | TAP | | | 749 | | |
| NAME: Cseri, Luann | | | QY | | | P | | | 900 | | |
| REGISTRATION NUMBER: 31,822 | | | Db | | | T | | | 809 | | |
| REFERENCE/DOCKET NUMBER: GI 5190 | | | QY | | | P | | | 960 | | |
| TELECOMMUNICATION INFORMATION: | | | Db | | | T | | | 869 | | |
| TELEPHONE: (617)876-1170 | | | QY | | | P | | | 1020 | | |
| TELEFAX: (617)876-5851 | | | Db | | | T | | | 929 | | |
| INFORMATION FOR SEQ ID NO: 74: | | | QY | | | P | | | 1080 | | |
| SEQUENCE CHARACTERISTICS: | | | Db | | | T | | | 989 | | |
| LENGTH: 1038 amino acids | | | QY | | | P | | | 1140 | | |
| TYPE: amino acid | | | Db | | | T | | | 1006 | | |
| TOPOLOGY: linear | | | QY | | | P | | | 1200 | | |
| MOLECULE TYPE: protein | | | Db | | | T | | | 1006 | | |
| SEQUENCE DESCRIPTION: SEQ ID NO: 74: | | | QY | | | P | | | 1260 | | |
| US-10-124-557-74 | | | Db | | | T | | | 1006 | | |
| Query Match | | | QY | | | P | | | 1361 | | |
| Best local Similarity 76.3%; Pred. No. 4.2e-125; Indels 323; Gaps 3; | | | Db | | | T | | | 1006 | | |
| Matches 1038; Conservative 0; Mismatches 0 | | | QY | | | P | | | 1320 | | |
| 1 MAWKTLPIYLLLLSVFVIQQVSSQLSSCAGRCGEGYSRDATCNCYDNCQHYNECCPDF 60 | | | Db | | | T | | | 1006 | | |
| 1 MAWKTLPIYLLLLSVFVIQQVSSQ-----25 | | | QY | | | P | | | 1200 | | |
| 61 KRVCUTBELSCKGRCFESFEREGRECDCAQCKKYDKCCPDYEFCAEVHNPTSPSSKKAP 120 | | | Db | | | T | | | 1006 | | |
| 26 -----ELSCKGRCFESFEREGRECDCAQCKKYDKCCPDYEFCA-----64 | | | QY | | | P | | | 1260 | | |
| 121 PPGASQTIKSTTKRSKPPNKKTKKVEIEITEVKNKKNRTKKKTPKPPVWDEAG 180 | | | Db | | | T | | | 1006 | | |
| 65 -----EVKNKKNRTKKKTPKPPVWDEAG 89 | | | QY | | | P | | | 1320 | | |
| 181 SGLDNGDPKVTTPDTSITQHNKVTSPKITTAKPINRPSPLPNSDTSKETSILVNKETT 240 | | | Db | | | T | | | 1006 | | |
| 90 SGLDNGDFKVTTPDTSITQHNKVTSPKITTAKPINRPSPLPNSDTSKETSILVNKETT 149 | | | QY | | | P | | | 1361 | | |
| 241 VETKETTNNKQSTGDKKETSAKETOSIEKTSAKOLAPTSKULAKETPKAEITTKGPA 300 | | | Db | | | T | | | 1006 | | |
| 150 VETKETTNNKQSTGDKKETSAKETOSIEKTSAKOLAPTSKULAKETPKAEITTKGPA 209 | | | QY | | | P | | | 1320 | | |
| 301 LTTTKEPTTTPKEPASTTPKEPTPTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 360 | | | Db | | | T | | | 1006 | | |
| 210 LTTTKEPTTTPKEPASTTPKEPTPTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 269 | | | QY | | | P | | | 1361 | | |
| 361 PAPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 420 | | | Db | | | T | | | 1006 | | |
| 270 PAPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 329 | | | QY | | | P | | | 1361 | | |
| 421 PKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480 | | | Db | | | T | | | 1006 | | |
| 330 PKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 389 | | | QY | | | P | | | 1361 | | |
| 481 PTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 540 | | | Db | | | T | | | 1006 | | |
| 390 PTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 449 | | | QY | | | P | | | 1361 | | |
| 541 PAPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 600 | | | Db | | | T | | | 1006 | | |
| 450 PAPTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 509 | | | QY | | | P | | | 1361 | | |
| 601 TPEELAPTTPEEPTPTTPEEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660 | | | Db | | | T | | | 1006 | | |
| 510 TPEELAPTTPEEPTPTTPEEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 569 | | | QY | | | P | | | 1361 | | |
| 661 TAPTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 720 | | | Db | | | T | | | 1006 | | |
| 570 TAPTPKGTAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 629 | | | QY | | | P | | | 1361 | | |
| 721 EPAPTPKGPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 780 | | | Db | | | T | | | 1006 | | |
| 630 EPAPTPKGPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 689 | | | QY | | | P | | | 1361 | | |

Search completed: October 13, 2004, 11:53:23
Job time : 146.466 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 149.642 Seconds
(without alignments)
5233.063 Million cell updates/sec

Title: SEQ1-G
Perfect score: 7323
Sequence: 1 MAWKTLPIYLLLSVEFIQ.....ARALITRSQTLSKVWVNC 1361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 7308.7 | 99.8 | 1404 | 2 Q92954 | Q92954 homo sapien |
| 2 | 7293.7 | 99.6 | 1404 | 2 Q9BX49 | Q9BX49 homo sapien |
| 3 | 4634.6 | 63.3 | 933 | 2 Q6ZM25 | Q6ZM25 homo sapien |
| 4 | 4634.6 | 63.3 | 933 | 2 BAD18580 | Bad18580 h cdna fl |
| 5 | 3395.3 | 54.6 | 1054 | 2 Q9UM99 | Q9UM99 mus musculus |
| 6 | 1714.7 | 23.4 | 401 | 2 Q77765 | Q77765 bos taurus |
| 7 | 1413.6 | 19.3 | 5179 | 1 MUC2 HUMAN | Q02817 homo sapien |
| 8 | 1381.2 | 18.9 | 1225 | 2 Q9VR49 | Q9VR49 drosophila |
| 9 | 1339.8 | 18.3 | 1761 | 2 Q7KTF6 | Q7KTF6 drosophila |
| 10 | 1339.8 | 18.3 | 1761 | 2 AAS64673 | AAS64673 drosophila |
| 11 | 1333 | 18.2 | 251 | 2 Q6DNC4 | Q6DNC4 homo sapien |
| 12 | 1310.1 | 17.9 | 3150 | 2 Q7PMD5 | Q7PMD5 anopheles g |
| 13 | 1256.8 | 17.2 | 3409 | 2 Q6SSE6 | Q6SSE6 chlamydomon |
| 14 | 1256.8 | 17.2 | 3409 | 2 AAS07044 | AAS07044 chlamydom |
| 15 | 1184.8 | 16.2 | 3889 | 2 Q6SSE8 | Q6SSE8 chlamydomon |
| 16 | 1184.8 | 16.2 | 3889 | 2 AAS07042 | AAS07042 chlamydom |
| 17 | 1183 | 16.2 | 251 | 2 Q6QJF3 | Q6QJF3 bos taurus |
| 18 | 1183 | 16.2 | 251 | 2 AAS83977 | AAS83977 bos taurus |
| 19 | 1179 | 16.1 | 1664 | 1 SLPI_CLOTM | Q06852 clostridium |
| 20 | 1139.6 | 15.6 | 5703 | 1 MUSE_HUMAN | Q9HC84 homo sapien |
| 21 | 1137.8 | 15.5 | 9234 | 2 Q7KTF5 | Q7KTF5 drosophila |
| 22 | 1137.8 | 15.5 | 9234 | 2 AAN10531 | AAN10531 drosophila |
| 23 | 1121.9 | 15.3 | 3432 | 2 Q8IR51 | Q8IR51 drosophila |
| 24 | 1121.8 | 15.3 | 3432 | 2 Q8IR52 | Q8IR52 drosophila |
| 25 | 1118.3 | 15.3 | 23015 | 2 Q8IQ18 | Q8IQ18 drosophila |
| 26 | 1118.3 | 15.3 | 23015 | 2 AAN10358 | AAN10358 drosophila |
| 27 | 1117 | 15.2 | 1795 | 2 Q76894 | Q76894 drosophila |
| 28 | 1115.4 | 15.2 | 1349 | 2 Q8WQ4 | Q8WQ44 homo sapien |
| 29 | 1108.3 | 15.1 | 1079 | 2 Q9N4S7 | Q9N4S7 caenorhabdi |
| 30 | 1096.9 | 15.0 | 34350 | 2 Q8WZ42 | Q8WZ42 homo sapien |
| 31 | 1080.9 | 14.8 | 2284 | 2 Q9VPG1 | Q9VPG1 drosophila |

ALIGNMENTS

RESULT 1

| | | | | |
|--------|--|--------------|------|----------|
| Q92954 | Q92954 | PRELIMINARY; | PRT; | 1404 AA. |
| AC | Q92954; | | | |
| DT | 01-FEB-1997 (TrEMBLrel. 02, Created) | | | |
| DT | 01-FEB-1997 (TrEMBLrel. 02, Last sequence update) | | | |
| DT | 01-MAR-2004 (TrEMBLrel. 26, Last annotation update) | | | |
| DE | Megakaryocyte stimulating factor. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C., | | | |
| RA | Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J., | | | |
| RA | Ferenz C., Grohholz J., Fraser H., Bean K., Norton C.R., Gesner T., | | | |
| RA | Bhatia S., Kriz R., Hewick R., Clark S.C.; | | | |
| RT | "Purification, Biochemical Characterization, and Cloning of a Novel | | | |
| RT | Megakaryocyte Stimulating Factor that has Megakaryocyte Colony | | | |
| RT | Stimulating Activity."; | | | |
| RL | Blood 78:279-279(1991). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P., | | | |
| RA | Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R., | | | |
| RA | Jacobs K., Turner K.; | | | |
| RT | "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor."; | | | |
| RL | (in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J., | | | |
| RL | Mosher D.F. (eds.); | | | |
| RL | BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier | | | |
| RL | Science Publishers B.V. (1993). | | | |
| RN | [3] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RA | Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C., | | | |
| RA | Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J., | | | |
| RA | Ferenz C., Grohholz J., Fraser H., Bean K., Norton C.R., Gesner T., | | | |
| RA | Bhatia S., Kriz R., Hewick R., Clark S.C.; | | | |
| RL | Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases. | | | |
| DR | EMBL; U70136; AAB09089.1; -; | | | |
| DR | HSSP; P04004; LOCO. | | | |
| DR | Genew; HGNC:9364; PRG4. | | | |
| DR | GO; GO:0008283; P:cell proliferation; TAS. | | | |
| DR | InterPro; IPR000585; Hemopexin. | | | |
| DR | InterPro; IPR001212; Somatomedin_B. | | | |
| DR | Pfam; PF00045; Hemopexin; 2. | | | |
| DR | Pfam; PF01033; Somatomedin_B; 2. | | | |
| DR | PRINTS; PR00022; SOMATOMEDINB. | | | |
| DR | SMART; SM00120; HX; 2. | | | |
| DR | SMART; SM00201; SO; 2. | | | |
| DR | PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1. | | | |
| DR | PROSITE; PS00524; SOMATOMEDIN_B; 2. | | | |
| SQ | SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64; | | | |

| | | | | | | |
|----|--------|------|-------|---|----------|---------------------|
| 32 | 1061.2 | 14.5 | 7962 | 2 | Q10465 | Q10465 homo sapien |
| 33 | 1050.9 | 14.4 | 2187 | 2 | P70670 | P70670 mus musculus |
| 34 | 1049 | 14.3 | 1607 | 2 | Q8H6Q5 | Q8H6Q5 phytophthor |
| 35 | 1049 | 14.3 | 1607 | 2 | AAP74661 | AAP74661 phytophth |
| 36 | 1036 | 14.1 | 972 | 2 | Q7OKK7 | Q7OKK7 anopheles g |
| 37 | 1027.3 | 14.0 | 1489 | 2 | Q96449 | Q96449 phytophthor |
| 38 | 1036.2 | 14.0 | 4498 | 2 | Q9W2Z3 | Q9W2Z3 drosophila |
| 39 | 1021.7 | 14.0 | 10578 | 2 | Q8ISF5 | Q8ISF5 caenorhabdi |
| 40 | 1021.6 | 14.0 | 1480 | 2 | Q9LIE8 | Q9LIE8 arabidopsis |
| 41 | 1021.1 | 13.9 | 18519 | 2 | Q8ISF6 | Q8ISF6 caenorhabdi |
| 42 | 1021.1 | 13.9 | 18534 | 2 | Q8ISF7 | Q8ISF7 caenorhabdi |
| 43 | 1018.4 | 13.9 | 1274 | 2 | Q20007 | Q20007 caenorhabdi |
| 44 | 1009.6 | 13.8 | 2768 | 2 | Q9VC00 | Q9VC00 drosophila |
| 45 | 1000.2 | 13.7 | 1458 | 2 | Q757N5 | Q757N5 ashbya goss |

| | | | | | | |
|-----------------------|-----|---|---------------------|------------|--------------|--------------------|
| Query Match | | 99.8%; | Score 7308.7; | DB 2; | Length 1404; | |
| Best Local Similarity | | 96.9%; | Pred. No. 5.6e-143; | | | |
| Matches 1361; | | Conservative | 0; | Mismatches | 0; | Indels 43; Gaps 1; |
| QY | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF | 60 | | | |
| Db | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF | 60 | | | |
| QY | 61 | KRVCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP | 120 | | | |
| Db | 61 | KRVCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP | 120 | | | |
| QY | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITE----- | 156 | | | |
| Db | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITE----- | 156 | | | |
| QY | 157 | -----VKDNKQNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDST | 197 | | | |
| Db | 181 | KIKSSKNSAANRELQKLVKDNKQNRKKKPKPPVVDVDEAGSLDNGDFKVTTPDST | 240 | | | |
| QY | 198 | TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVEHKEITTTNKQTSIDG | 257 | | | |
| Db | 241 | TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVEHKEITTTNKQTSIDG | 300 | | | |
| QY | 258 | KEKTSKAKTOSIEKTSKDLAPTSKVLAKPTKAEITTKGPALTTTPKEPTPTTKBPAS | 317 | | | |
| Db | 301 | KEKTSKAKTOSIEKTSKDLAPTSKVLAKPTKAEITTKGPALTTTPKEPTPTTKBPAS | 360 | | | |
| QY | 318 | TTPKPEPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTKPAPTTTKP | 377 | | | |
| Db | 361 | TTPKPEPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTKPAPTTTKP | 420 | | | |
| QY | 378 | APTTPKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTK | 437 | | | |
| Db | 421 | APTTPKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTK | 480 | | | |
| QY | 438 | EPAPTAEPKAPPTTPKEPAPTTTPKEPAPTTTPKESPTTPKEPAPTTTKSAPTTTPKAPT | 497 | | | |
| Db | 481 | EPAPTAEPKAPPTTPKEPAPTTTPKEPAPTTTPKESPTTPKEPAPTTTKSAPTTTPKAPT | 540 | | | |
| QY | 498 | TTKSAPTTTPKESPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTKKP | 557 | | | |
| Db | 541 | TTKSAPTTTPKESPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTKKP | 600 | | | |
| QY | 558 | APTAPKBPAPTTTPKETAPTTPKKLTTPTPKEKLAPTTTPKEAPTTPEELAPTTPEPTPT | 617 | | | |
| Db | 601 | APTAPKBPAPTTTPKETAPTTPKKLTTPTPKEKLAPTTTPKEAPTTPEELAPTTPEPTPT | 660 | | | |
| QY | 618 | PEEPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP | 677 | | | |
| Db | 661 | PEEPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP | 720 | | | |
| QY | 678 | APTTPKPKAPKELAPTTTKEPTSTSDKAPAPTTTPKGTAPTTPKAPAPTTTPKEPAPTTPKG | 737 | | | |
| Db | 721 | APTTPKPKAPKELAPTTTKEPTSTSDKAPAPTTTPKGTAPTTPKAPAPTTTPKEPAPTTPKG | 780 | | | |
| QY | 738 | TAPTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPAPTTTPKETAPTTPKEPAPTTPK | 797 | | | |
| Db | 781 | TAPTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPAPTTTPKETAPTTPKEPAPTTPK | 840 | | | |
| QY | 798 | KPAPTTPETPPTTSEVSTPTTTPKEPTTIHKSPDESTPELSAETTPKALENSPKPGVPT | 857 | | | |
| Db | 841 | KPAPTTPETPPTTSEVSTPTTTPKEPTTIHKSPDESTPELSAETTPKALENSPKPGVPT | 900 | | | |
| QY | 858 | TKTPAATKPEMTTAKOKTTERDLRTPPETTTAAAPKMTKETATTTTEKTTESKITATTQV | 917 | | | |
| Db | 901 | TKTPAATKPEMTTAKOKTTERDLRTPPETTTAAAPKMTKETATTTTEKTTESKITATTQV | 960 | | | |
| QY | 918 | TSTTTQDTPPKITLTKTLTKTLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTPK | 977 | | | |
| Db | 961 | TSTTTQDTPPKITLTKTLTKTLAPKVTITTKITITTEIMNKPEETAKPKDRATNSKATTPK | 1020 | | | |
| QY | 978 | PQKPTKAPKPTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN | 1037 | | | |

| | | | | | | |
|-----------------------|---|--|---------------|------------|--------------|--------------------|
| Db | 1021 | PQKPTKAPKPTSTTKPKTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN | 1080 | | | |
| QY | 1038 | QTPNSKLVNPKSESDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS | 1097 | | | |
| Db | 1081 | QTPNSKLVNPKSESDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS | 1140 | | | |
| QY | 1098 | DETNICNGKPVLDGLTTLRNGTLVAFRGHYFWMLSPEFSPARRITEVMGIPSPIDTFT | 1157 | | | |
| Db | 1141 | DETNICNGKPVLDGLTTLRNGTLVAFRGHYFWMLSPEFSPARRITEVMGIPSPIDTFT | 1200 | | | |
| QY | 1158 | RCNCEGKTPFFKDSQWRFNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPESVY | 1217 | | | |
| Db | 1201 | RCNCEGKTPFFKDSQWRFNDIKDAGYKPKIPKFGGLTGQIVAAALSTAKYKNWPESVY | 1260 | | | |
| QY | 1218 | FFKRGSSIQQYIYKQBPVQKCPGRPALNYPVYGMTQVRRRRFERAIGFSQTHTRIQQ | 1277 | | | |
| Db | 1261 | FFKRGSSIQQYIYKQBPVQKCPGRPALNYPVYGMTQVRRRRFERAIGFSQTHTRIQQ | 1320 | | | |
| QY | 1278 | SPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKQYYNIDV | 1337 | | | |
| Db | 1321 | SPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKQYYNIDV | 1380 | | | |
| QY | 1338 | PSRTARAITTRSGQTLISKVWYNCP | 1361 | | | |
| Db | 1381 | PSRTARAITTRSGQTLISKVWYNCP | 1404 | | | |
| RESULT 2 | | | | | | |
| Q9BX49 | | | | | | |
| ID | Q9BX49 | PRELIMINARY; | PRF; | 1404 | AA. | |
| AC | Q9BX49; | | | | | |
| DT | 01-JUN-2001 | (T=EMBLrel. 17, Created) | | | | |
| DT | 01-JUN-2001 | (T=EMBLrel. 17, Last sequence update) | | | | |
| DT | 01-MAR-2004 | (T=EMBLrel. 26, Last annotation update) | | | | |
| DE | BGI7416.2 | (MSF: megakaryocyte stimulating factor) | | | | |
| GN | Name=BG174L6.2; | | | | | |
| OS | Homo sapiens (Human) | | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | | |
| OX | NCBI_TaxID=9606; | | | | | |
| RN | [1] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RA | Wray P.; | | | | | |
| RL | Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases. | | | | | |
| DR | EMBL; AL133553; CAC36090.1; - | | | | | |
| DR | HSSP; P04004; 10C0. | | | | | |
| DR | InterPro; IPR000585; Hemopexin. | | | | | |
| DR | InterPro; IPR001212; Somatomedin_B. | | | | | |
| DR | Pfam; PF00045; Hemopexin; 2. | | | | | |
| DR | Pfam; PF01033; Somatomedin_B; 2. | | | | | |
| DR | PRINTS; PR00022; SOMATOMEDINB. | | | | | |
| DR | SMART; SM00120; HX; 2. | | | | | |
| DR | SMART; SM00201; SO; 2. | | | | | |
| DR | PROSITE; PS00024; HMOPEXIN; UNKNOWN 1. | | | | | |
| DR | PROSITE; PS00524; SOMATOMEDIN B; 2. | | | | | |
| SQ | SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64; | | | | | |
| Query Match | | | | | | |
| Best Local Similarity | | 99.6%; | Score 7293.7; | DB 2; | Length 1404; | |
| Matches 1358; | | Conservative | 0; | Mismatches | 3; | Indels 43; Gaps 1; |
| QY | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF | 60 | | | |
| Db | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF | 60 | | | |
| QY | 61 | KRVCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP | 120 | | | |
| Db | 61 | KRVCTAELSCKRCFCFESFERGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP | 120 | | | |
| QY | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITE----- | 156 | | | |
| Db | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITE----- | 156 | | | |

QY 157 -----VKDNKKNRITKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 197
 Db 181 KIKSSKNSAANRELQKKLVKDNKKNRITKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
 QY 198 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKOTSTG 257
 Db 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKOTSTG 300
 QY 258 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGALTPKBPPTTTPKEPAS 317
 Db 301 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGALTPKBPPTTTPKEPAS 360
 QY 318 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTPKEPAPTTTKEP 377
 Db 361 TTPKEPTTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTPKEPAPTTTKEP 420
 QY 378 APITTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 437
 Db 421 APITTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480
 QY 438 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 497
 Db 481 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 540
 QY 498 TTKGAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 557
 Db 541 TTKGAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 600
 QY 558 APITAPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 617
 Db 601 APITAPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 660
 QY 618 PEEBPAPTTTPKAAADNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 677
 Db 661 PEEBPAPTTTPKAAADNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 720
 QY 678 APITPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 737
 Db 721 APITPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 780
 QY 738 TAPITLKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 797
 Db 781 TAPITLKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPT 840
 QY 798 KPAPTTPETPPPTTSEVSTPTTTPKEPTTIHKSDESTPELSABPTPKALENSKPEGVPT 857
 Db 841 KPAPTTPETPPPTTSEVSTPTTTPKEPTTIHKSDESTPELSABPTPKALENSKPEGVPT 900
 QY 858 TKTPAATKPEMTTAKDKTTERDLRTTTPETTTAAPKMTKETATTTTEKTTESKITATTTOV 917
 Db 901 TKTPAATKPEMTTAKDKTTERDLRTTTPETTTAAPKMTKETATTTTEKTTESKITATTTOV 960
 QY 918 TSTTTTODTTPKEKITTLTKTTTLAPKVTITTKTIITTEIMNKPEBTAKPORATNSKATTPK 977
 Db 961 TSTTTTODTTPKEKITTLTKTTTLAPKVTITTKTIITTEIMNKPEBTAKPORATNSKATTPK 1020
 QY 978 PQKPTKAPKKPTSTKKPKTMVRVRPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRPN 1037
 Db 1021 PQKPTKAPKKPTSTKKPKTMVRVRPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRPN 1080
 QY 1038 QTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 1097
 Db 1081 QTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIIINPMLS 1140
 QY 1098 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPARRITTEVWGIFSPIDTFT 1157
 Db 1141 DETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPARRITTEVWGIFSPIDTFT 1200
 QY 1158 RCNCEGKTPFFFKDSQWRFTNDIKDAGYKPKIFKGGGLTGQIVAAALSTAKYKNWPESVY 1217
 Db 1201 RCNCEGKTPFFFKDSQWRFTNDIKDAGYKPKIFKGGGLTGQIVAAALSTAKYKNWPESVY 1260

QY 1218 FPKRGGSIQQYIYKQEPVQKCFRRPALNPYVYGMTQVRRRRFERAIGPSOHTHRIQY 1277
 Db 1261 FPKRGGSIQQYIYKQEPVQKCFRRPALNPYVYGMTQVRRRRFERAIGPSOHTHRIQY 1320
 QY 1278 SPARLAYODKGVHLNEVKVSLIWRGLPNVVTSAISLNPTRKPDGYDYAFSKDQYNNIDV 1337
 Db 1321 SPARLAYODKGVHLNEVKVSLIWRGLPNVVTSAISLNPTRKPDGYDYAFSKDQYNNIDV 1380
 QY 1338 PSRTARAITTRSGQTLKSVWYNCP 1361
 Db 1381 PSRTARAITTRSGQTLKSVWYNCP 1404
 RESULT 3
 Q6ZMZ5
 ID Q6ZMZ5 PRELIMINARY; PRT; 933 AA.
 AC Q6ZMZ5;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE Hypothetical protein FLJ16561.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Synovial membrane tissue;
 RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
 RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Sugiyama M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
 RA Masuho Y., Nagai K., Isegai T.,
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK131434; BAB18580.1; -;
 DR InterPro; IPR000585; Hemopexin.
 DR InterPro; IPR001212; Somatomedin B.
 DR Pfam; PF00045; Hemopexin; 2.
 DR Pfam; PF01033; Somatomedin B; 1.
 DR PRINTS; PR00022; SOMATOMEDINB.
 DR SMART; SM00120; HX; 2.
 DR SMART; SM00201; SO; 1.
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
 DR PROSITE; PS00524; SOMATOMEDIN B; 1.
 SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;
 Query Match 63.3%; Score 4634.6; DB 2; Length 933;
 Best Local Similarity 63.3%; Pred. No. 6.5e-88;
 Matches 889; Conservative 0; Mismatches 1; Indels 514; Gaps 3;
 QY 1 MAWKTLPIYLLLLLSVFIQQVSSQDLSSCAGRCGEGYSDATCNDYNCQHYMECCPDF 60
 Db 1 MAWKTLPIYLLLLLSVFIQQVSSQ----- 25
 QY 61 KRVTAEUSCKRCFESFERGECDCDAOCKKYDKCCPDYESFCAEVHNPTSPBSKKAP 120
 Db 26 -----ELSCKRCFESFERGECDCDAOCKKYDKCCPDYESFCAEVHNPTSPBSKKAP 79
 QY 121 PPSGASQITKSTTKRSPKPPNKKTKKVIESBEITE----- 156
 Db 80 PPSGASQITKSTTKRSPKPPNKKTKKVIESBEITEHSVSNQESSSSSSSSSSSTIR 139
 QY 157 -----VKDNKKNRITKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 197
 Db 140 KIKSSKNSAANRELQKKLVKDNKKNRITKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 199
 QY 198 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKOTSTG 257
 Db 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTNNKOTSTG 259
 QY 258 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGALTPKBPPTTTPKEPAS 317

Db 910 PSRTARAITRRSQGLTSKVWYNCP 933

RESULT 4
BAD18580 PRELIMINARY; PRT; 933 AA.

ID BAD18580
AC BAD18580;
DT 12-MAY-2004 (T-EMBLrel. 27, Created)
DT 12-MAY-2004 (T-EMBLrel. 27, Last sequence update)
DT 12-MAY-2004 (T-EMBLrel. 27, Last annotation update)
DE CNA FLJ16561 fis, clone SYN04003981, moderately similar to Homo sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular superficial zone protein, (megakaryocyte stimulating factor, articular pericarditis syndrome) (PRG4)).
DE OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma A., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BAD18580.1, ..
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 63.3%; Score 4634.6; DB 2; Length 933;
Best Local Similarity 63.3%; Pred. No. 6.5e-88;
Matches 889; Conservative 0; Mismatches 1; Indels 514; Gaps 3;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
QY 61 KEVCTAELSCKGRCPESPERGREGDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
DB 26 -----EUSCKGRCPESPERGREGDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
QY 121 PPSGASQTIKSTTKESPKPPNKKTKKVIIESEITE----- 156
DB 80 PPSGASQTIKSTTKESPKPPNKKTKKVIIESEITEEHSVSESSSSSSSSSSSSTIR 139
QY 157 -----VKDNKNRTKKKTPKPPVVVDAGSLDNGDKPVTPDTST 197
DB 140 KIKSSKNSAANRELQKLVKDNKNRTKKKTPKPPVVVDAGSLDNGDKFVITPDIST 199
QY 198 TQHNKVSPTSPLITAKPINRPSLPNSDTSKETSITVKNKETTETTTNNKQTSIDG 257
DB 200 TQHNKVSPTSPLITAKPINRPSLPNSDTSKETSITVKNKETTETTTNNKQTSIDG 259
QY 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAAPTSKVAETTTKGPALITPKETPTTKEPAS 317
DB 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAAPTSKVAETTTKGPALITPKETPTTKEPAS 319
QY 318 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 377
DB 320 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKE 370
QY 378 APITTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 437
DB 371 ----- 370
QY 438 EPAPTAPKPAFTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 497
DB 371 ----- 370

260 KEKTTSAKETQSIEKTSKADLAPTSKVLAAPTSKVAETTTKGPALITPKETPTTKEPAS 319
QY 318 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEP 377
DB 320 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKE 370
QY 378 APITTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 437
DB 371 ----- 370
QY 438 EPAPTAPKPAFTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 497
DB 371 ----- 370
QY 498 TTKSAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 557
DB 371 ----- 370
QY 558 APAPTAPKPAFTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 617
DB 371 ----- 370
QY 618 PEEPAPTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT 677
DB 371 ----- 370
QY 678 APPTPKKPAKELAPTTTKEPTSTTSKPAFTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKG 737
DB 371 ----- 370
QY 738 TAPPTLKEPAPTTTPKPAKELAPTTTKPTSTTSKPAFTTPKETATTPKEPAPTTTPK 797
DB 371 ----- 370
QY 798 KPAPTTPPTTSEVSTPTTKBPTTIHKSPDESTPELSAETTPKALENPKPGVPT 857
DB 371 -PAPTTPPTTSEVSTPTTKBPTTIHKSPDESTPELSAETTPKALENPKPGVPT 429
QY 858 TKTPAATKDEMTTAKDKTTERDLRTPPTTAAAPKMKETATTTKTESKITATTQV 917
DB 430 TKTPAATKDEMTTAKDKTTERDLRTPPTTAAAPKMKETATTTKTESKITATTQV 489
QY 918 TSTTTQDTTPFKITLTKTTTLAPKVTITTKKTTTTIMNKPETAKPKDRATNSKATTPK 977
DB 490 TSTTTQDTTPFKITLTKTTTLAPKVTITTKKTTTTIMNKPETAKPKDRATNSKATTPK 549
QY 978 POKPTKAPKPTSTKPKTTPVRKPKTTPTPRKMSTTMPELNPTSRIAEAMLQTTTRN 1037
DB 550 POKPTKAPKPTSTKPKTTPVRKPKTTPTPRKMSTTMPELNPTSRIAEAMLQTTTRN 609
QY 1038 QTPNSKLVEVNPKSADAGAGETPHMLRPHVFMPEVPTDMDYLPVSNQGLINPMLIS 1097
DB 610 QTPNSKLVEVNPKSADAGAGETPHMLRPHVFMPEVPTDMDYLPVSNQGLINPMLIS 669
QY 1098 DETN1CNGKPDVGLTILRNGLTVAFGHVFWMLSPESSPARRITFVWGIPSPIDTVFT 1157
DB 670 DETN1CNGKPDVGLTILRNGLTVAFGHVFWMLSPESSPARRITFVWGIPSPIDTVFT 729
QY 1158 RCNCEGTFPFKDSQWRFNTDKDAGYKPKIPKFGGGLTGQIVAAIATAKYKNWPEVY 1217
DB 730 RCNCEGTFPFKDSQWRFNTDKDAGYKPKIPKFGGGLTGQIVAAIATAKYKNWPEVY 789
QY 1218 FFKRGGSIQYIYKQBPVKCPGRPALNYPVNGEQTQVRRRRFERAIGPSQHTTIRIQY 1277
DB 790 FFKRGGSIQYIYKQBPVKCPGRPALNYPVNGEQTQVRRRRFERAIGPSQHTTIRIQY 849
QY 1278 SPARLAYQDKGVLHNEVKVSTLWRGLPNVNTSAISLFPNIRKPDGYDYAFSKQYYNIDV 1337
DB 850 SPARLAYQDKGVLHNEVKVSTLWRGLPNVNTSAISLFPNIRKPDGYDYAFSKQYYNIDV 909
QY 1338 PSRTARAITRRSQGLTSKVWYNCP 1361

```

QY 498 TTKSAPTTKBPSPPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 557
Db 371 ----- 370
QY 558 APTAPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTT 617
Db 371 ----- 370
QY 618 PEPAPTTKAAAPNTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 677
Db 371 ----- 370
QY 678 APPTPKPAPKELAPTTKBPSTTSDDKAPTTKBPAPTTKBPAPTTKBPAPTTK 737
Db 371 ----- 370
QY 738 TAPTTKBPAPTTKBPAPKELAPTTKBPSTTSDDKAPTTKBPAPTTKBPAPTTK 797
Db 371 ----- 370
QY 798 KPAPTTPTPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPPTKALENSKPE 857
Db 371 -PAPTTPTPTPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPPTKALENSKPE 429
QY 858 TKTPAAATKPEMTTAKOKTTERDRTTPTTAAAPKMTKETAATTTKTESKITATTQ 917
Db 430 TKTPAAATKPEMTTAKOKTTERDRTTPTTAAAPKMTKETAATTTKTESKITATTQ 489
QY 918 TSTTTQDTPFKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKAT 977
Db 490 TSTTTQDTPFKITTLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKAT 549
QY 978 PQKPTKAPKPTSTKPKPTMPVRKPKTTPPRKWTSTMPKLNPTSLIAEAMLOTT 1037
Db 550 PQKPTKAPKPTSTKPKPTMPVRKPKTTPPRKWTSTMPKLNPTSLIAEAMLOTT 609
QY 1038 QTNSKLVEVNPKSDAGGAETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINP 1097
Db 610 QTNSKLVEVNPKSDAGGAETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINP 669
QY 1098 DETNINCGKPVGGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITVWGIPI 1157
Db 670 DETNINCGKPVGGLTTLRNGTLVAFRGHYFWMLSPSPSPARRITVWGIPI 729
QY 1158 RCNCEGKTEFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIYVAALSTAKY 1217
Db 730 RCNCEGKTEFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIYVAALSTAKY 789
QY 1218 FPKRGSSIQQYIYKQBPVOKPGRRRPALNYPVYGEMTQVRRRRFERAIGPS 1277
Db 790 FPKRGSSIQQYIYKQBPVOKPGRRRPALNYPVYGEMTQVRRRRFERAIGPS 849
QY 1278 SPARLAYQDQVLHNEVKYSILWRGLPNVVTSAISLPNTRKPDGYDYAFSPK 1337
Db 850 SPARLAYQDQVLHNEVKYSILWRGLPNVVTSAISLPNTRKPDGYDYAFSPK 909
QY 1338 PSRTARAITRRSGQTLISKVYWNCP 1361
Db 910 PSRTARAITRRSGQTLISKVYWNCP 933

RESULT 5
Q9JUM99
ID Q9JUM99 PRELIMINARY; PRT; 1054 AA.
AC Q9JUM99;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus.
GN Name=Prg4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";
RL CytoGenet. Cell Genet. 90:291-297(2000).
DR EMBL; AB034730; BAA92310.1; -.
DR HSSP; P04004; 10C0.
DR MGD; MGI:1891344; Prg4.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR PRINTS; PF01033; Somatomedin_B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match
Best Local Similarity 54.6%; Score 3995.3; DB 2; Length 1054;
Matches 773; Conservative 58; Mismatches 163; Indels 427; Gaps 22;

QY 1 MAWKTLPIYLLLSVFVIOQVSSQDLSSCAGRCGEGYSDATCNCVNCOHYMECCPDF 60
Db 1 MGWKILPVCLSLLPVLIQVSSQDLSSCAGRCGEGYSDATCNCVNCOHYMECCPDF 60
QY 61 KRVCITABLSKGRCFESFERGRCDCDAQCKYDKCCPYESFCAEVHNPTSPSSKKAP 120
Db 61 KRVCSPELSKGRCFESFARGRCDCDSQCKYKCCADYDSFCEEVHNSTSPSSKAP 119
QY 121 PPGCASQTIKSTTKRSPKPNKKTKVIESEITE----- 156
Db 120 TPAGASDTIKSTTKRSPKPT-TRTIKWSESEITEHSNSENQESSSSSSSTIRKI 178
QY 157 -----VKDNKNRTKKPTPKPPVVDVDEAGSLDNGDFKVT--TPDTSTT 198
Db 179 KSSKNSANRELQKNPNVKNKNTPKKKNPEPAVDVDEAGSLDNGDFKUTPPDPDPPT 238
QY 199 QHNKVTSPKITTAKPINRPSLPNSDTSKETSLSLVNKETTVEKTTTNKQTSIDGK 258
Db 239 PHSKVATSPKTTAAKPVTPKPSLAPNSETSKESLASNKETTVEKTTATNKQSSA-SK 297
QY 259 EKTTSKAKETQSIEKTSKDLAPTSKVLAKPTPKAETTKGPAITTKPEPTTPPKBP 318
Db 298 KKTTSVKETRSABKTSKDOVEPTS----- 321
QY 319 TPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKBPAPTT-----PKEPA 370
Db 322 -----TPKNSAPTTTKKPVTTKSKFLPLPOEPE 352
QY 371 PTTTKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 430
Db 353 PTTAKEPPTTKKPEPTTKPEPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 412
QY 431 PAPTTTPKEPAPTPAKKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 490
Db 413 PEPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 461
QY 491 TKPEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 550
Db 462 -----EPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 510
QY 551 PTTTTPKEPAPTPAKKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 610
Db 511 PT----- 512
QY 611 BEPTPTTPEEPAPTTPKAAAPNTPKBPAPTTTPKEPAPTTTPKEPAPTTTPKE 670
Db 513 -----TPKEPPTTPKEPPTTPKEPPTTP----- 538

```

Biochem. Biophys. Res. Commun. 254:535-541(1999).

[2]

RL
RN
RP
SEQUENCE FROM N.A.
RC
TISSUE=Articular cartilage;
RX
MEDLINE=99171663; PubMed=10073655;
RA
Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
AYdelotte M.B.;
RA
"Immunodetection and partial cDNA sequence of the proteoglycan,
RT
superficial zone protein, synthesized by cells lining synovial
RT
joints.";
RL
J. Orthop. Res. 17:110-120(1999).
DR
EMBL; AF056218; AAD13404.1; -;
DR
InterPro; IPR000585; Hemoexin.
DR
Pfam; PF00045; Hemoexin; 2.
DR
SMART; SMO0120; HX; 2.
DR
PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
FT
NON_TER
FT
SEQUENCE 401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;
Query Match 23.4%; Score 1714.7; DB 2; Length 401;
Best Local Similarity 80.2%; Pred. No. 4.1e-28; Indels 3; Gaps 3;
Matches 320; Conservative 24; Mismatches 52;
QY 964 PKDRATNSKATTPKPKQKTKAPKPKTSTKPKTPRVRKPKTTPRKR-MTSTMPENLPT 1022
Db 5 PKGRATNSQVTTPKQKTKAPKPKTSTKPKTPRVRKPKTTPRKR-MTSTMPENLPT 63
QY 1023 SRIAEMLOTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMYL 1082
Db 64 S-LPEAMLOTTTRPTTNSSEIIVNSENEDGAAEGEKPHMIFRPPVLTPIVPGTEII 122
QY 1083 PRVFNQGIINPMLSDETNI CNKRPVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARRI 1142
Db 123 VRGSPQGGINPMFSDETNL CNKRPVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARRI 182
QY 1143 TEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 1202
Db 183 TEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 242
QY 1203 ALSTAKYKNWPSVYFFKRGSGIQIYIKQBEVQKCPGRRPALNYPVYGVEMTQVRRRRF 1262
Db 243 ALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTQCTGRFPAINYSVYGEAAQVRRRF 302
QY 1263 RAIGPSQTHIRIQYS-PARLAYODKGLVHNEVKVSLMRGLNPNVTSALSINIRKPD 1320
Db 303 RAIGPSQTHIRIHYSPVPMRVSQYQDKGFLHNEVKVSLMRGLNPNVTSALSINIRKPD 362
QY 1321 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 363 DYALSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 401
RESULT 7
MUC2_HUMAN STANDARD; PRT; 5179 AA.
ID MUC2_HUMAN
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
QY 671 PPTLKBPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTPKPEP 730
Db 539 -----KEPEPTTPKPP-----EPTTKEPVPTTPKPEP 565
QY 731 APPTPKGTAPTTLKEPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTPKPEP 790
Db 566 EPTTP-----KEPEPTTPKPEP-----TTRKEPEPTTPKPEPTTPKPEP 604
QY 791 PAPTTPKKPAPTPPTPTTSEVSTPTTKEPTTHKGPDESTPLSAPTPKALENSP 850
Db 605 PEPTTPKKPEPTT-----617
QY 851 KPEQVPTTKTAAKPEMTTAKDKITERDLRTTBETTTAAPKWTKETATITTEKTTESKI 910
Db 618 -----TSPKT-----622
QY 911 TATTQTQVSTTTQDTPFKITLTKTTLAPKVVTTTKTITTTIMNKPSTETAKPKDRATN 970
Db 623 -----TTLKATTLAPKVTAPAE-----EIQNKPEETTPASEDSDD 657
QY 971 SKAT-----TPKP-QKPTKAPKPTSTKPKTPRVRKPKTTPRKRMTSTMPENLAP 1021
Db 658 SKTTLKPKQPKAPKPKTKPKKPTSTKPKT-PKTRKPKTTPAPLTKTSATPELNT 716
QY 1022 TSRIAEMLOTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMOY 1081
Db 717 TP--LEVMLPTTITKQTPNPETAENVDPDHDADGGEKXP-LIPGPPVLFPPTAIPGTDL 773
QY 1082 LPRVFNQGIINPMLSDETNI CNKRPVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARR 1141
Db 774 LAGLRNIGINPSPDETNI CNKRPVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARR 833
QY 1142 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 1201
Db 834 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 893
QY 1202 AALSTAKYKNWPSVYFFKRGSGIQIYIKQBEVQKCPGRRPALNYPVYGVEMTQVRRRRF 1261
Db 894 AALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTQCTGRFPAINYSVYGEAAQVRRRF 953
QY 1262 ERAIGPSQTHIRIQYS-PARLAYODKGLVHNEVKVSLMRGLNPNVTSALSINIRKPD 1320
Db 954 ERAVGPFQTHIRIHYSPVPMRVSQYQDKGFLHNEVKVSLMRGLNPNVTSALSINIRKPD 1013
QY 1321 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 1014 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1054
RESULT 6
O77765 PRELIMINARY; PRT; 401 AA.
ID O77765
AC O77765;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Superficial zone protein (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Ruminantia; Bovidae.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99120896; PubMed=9920774;
RA Flannery C.R., Hughes C.E., Schumacher B.L., Tudor D., Aydelotte M.B.,
KUettner K.E., Caterson B.;
RT "Articular cartilage superficial zone protein (SZP) is homologous to
RT megakaryocyte stimulating factor precursor and is a multifunctional
RT proteoglycan with potential growth-promoting, cytoprotective, and
RT lubricating properties in cartilage metabolism.";
QY 671 PPTLKBPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTPKPEP 730
Db 539 -----KEPEPTTPKPP-----EPTTKEPVPTTPKPEP 565
QY 731 APPTPKGTAPTTLKEPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTPKPEP 790
Db 566 EPTTP-----KEPEPTTPKPEP-----TTRKEPEPTTPKPEPTTPKPEP 604
QY 791 PAPTTPKKPAPTPPTPTTSEVSTPTTKEPTTHKGPDESTPLSAPTPKALENSP 850
Db 605 PEPTTPKKPEPTT-----617
QY 851 KPEQVPTTKTAAKPEMTTAKDKITERDLRTTBETTTAAPKWTKETATITTEKTTESKI 910
Db 618 -----TSPKT-----622
QY 911 TATTQTQVSTTTQDTPFKITLTKTTLAPKVVTTTKTITTTIMNKPSTETAKPKDRATN 970
Db 623 -----TTLKATTLAPKVTAPAE-----EIQNKPEETTPASEDSDD 657
QY 971 SKAT-----TPKP-QKPTKAPKPTSTKPKTPRVRKPKTTPRKRMTSTMPENLAP 1021
Db 658 SKTTLKPKQPKAPKPKTKPKKPTSTKPKT-PKTRKPKTTPAPLTKTSATPELNT 716
QY 1022 TSRIAEMLOTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMOY 1081
Db 717 TP--LEVMLPTTITKQTPNPETAENVDPDHDADGGEKXP-LIPGPPVLFPPTAIPGTDL 773
QY 1082 LPRVFNQGIINPMLSDETNI CNKRPVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARR 1141
Db 774 LAGLRNIGINPSPDETNI CNKRPVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARR 833
QY 1142 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 1201
Db 834 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 893
QY 1202 AALSTAKYKNWPSVYFFKRGSGIQIYIKQBEVQKCPGRRPALNYPVYGVEMTQVRRRRF 1261
Db 894 AALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTQCTGRFPAINYSVYGEAAQVRRRF 953
QY 1262 ERAIGPSQTHIRIQYS-PARLAYODKGLVHNEVKVSLMRGLNPNVTSALSINIRKPD 1320
Db 954 ERAVGPFQTHIRIHYSPVPMRVSQYQDKGFLHNEVKVSLMRGLNPNVTSALSINIRKPD 1013
QY 1321 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 1014 GYDYAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1054

Biochem. Biophys. Res. Commun. 254:535-541(1999).

[2]

RL
RN
RP
SEQUENCE FROM N.A.
RC
TISSUE=Articular cartilage;
RX
MEDLINE=99171663; PubMed=10073655;
RA
Schumacher B.L., Hughes C.E., Kuettnner K.E., Caterson B.,
AYdelotte M.B.;
RA
"Immunodetection and partial cDNA sequence of the proteoglycan,
RT
superficial zone protein, synthesized by cells lining synovial
RT
joints.";
RL
J. Orthop. Res. 17:110-120(1999).
DR
EMBL; AF056218; AAD13404.1; -;
DR
InterPro; IPR000585; Hemoexin.
DR
Pfam; PF00045; Hemoexin; 2.
DR
SMART; SMO0120; HX; 2.
DR
PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
FT
NON_TER
FT
SEQUENCE 401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;
Query Match 23.4%; Score 1714.7; DB 2; Length 401;
Best Local Similarity 80.2%; Pred. No. 4.1e-28;
Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;
QY 964 PKDRATNSKATTPKPKQKTKAPKPKTSTKPKTPRVRKPKTTPRKR-MTSTMPENLPT 1022
Db 5 PKGRATNSQVTTPKQKTKAPKPKTSTKPKTPRVRKPKTTPRKR-MTSTMPENLPT 63
QY 1023 SRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMYL 1082
Db 64 S-LPEAMLQTTTRPTPTNSEIIVNSENEDGAAEGEKPHMIFRPPVLTPIVPGTEII 122
QY 1083 PRVFNQGIILNPLSDETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMFLSPSPSPARRI 1142
Db 123 VRGSPQGGINPMFSDETNLNCGRPVDGLTLRLNGTLVAFRGHYFWMFLTPPTPPPPRRI 182
QY 1143 TEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIVA 1202
Db 183 TEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIVA 242
QY 1203 ALSTAKYKNWPSVYFFKRGSGIQYIYKQBEVQKCPGRRPALNYPVYGVEMTQVRRRRF 1262
Db 243 ALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTQCTGRFPAINYSVYGETAQVRRRRF 302
QY 1263 RAIGPSQTHIRIQYS-PARLAYQDKGLVHNEVKVSLMRGLNPNVTSALSINIRKPDGY 1322
Db 303 RAIGPSQVHTIRIHYPVPRVYQDKGFLHNEVKVSLMRGLNPNVTSALSINIRKPDGY 362
QY 1323 DYVAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 363 DYVAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 401
RESULT 7
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
QY 671 PPTLKBPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTPPKPEP 730
Db 539 -----KEPEPTTPKKP-----EPTTKEPVPTPKPEP 565
QY 731 APPTPKGTAPTTLKEPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPPKETAPTPKPE 790
Db 566 EPTTP-----KEPEPTTPKKPEP-----TTRKEPEPTTPKEPEPTPKPE 604
QY 791 PAPTTPKKPAPTTPETPTTSEVSTPTTKEPTTHKGPDESTPLSAPETPKALENSP 850
Db 605 PEPTTPKKPEPTT-----617
QY 851 KPEQVPTTKTAAKPEMTTAKDKITERDLRTTBETTTAAPKWTKETATITTEKTTESKI 910
Db 618 -----TSPKT-----622
QY 911 TATTQTQVSTTTQDTPFKITLTKTTLAPKVVTTTKTITTTIMNKPSTETAKPKDRATN 970
Db 623 -----TTLKATTLAPKVTAPAE-----EIQNKPEETTPASEDSDD 657
QY 971 SKAT-----TPKP-QKPTKAPKKTSTKPKTPRVRKPKTTPRKRMTSTMPENLAP 1021
Db 658 SKTTLKPKQPKAPKPKTKPKKPTSTKPKT-PKTRKPKTTPAPLTKTSATPELNT 716
QY 1022 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMOY 1081
Db 717 TP--LEVMLPFTTIPKQPNPETAENVPDHEDADGEGEKP-LIPGPPVLFPPTAIPGTDL 773
QY 1082 LPRVFNQGIILNPLSDETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMFLSPSPSPARR 1141
Db 774 LAGRLNRGININPSPDETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMFLSPSPSPARR 833
QY 1142 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 1201
Db 834 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 893
QY 1202 AALSTAKYKNWPSVYFFKRGSGIQYIYKQBEVQKCPGRRPALNYPVYGVEMTQVRRRRF 1261
Db 894 AALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTQCTGRFPAINYSVYGETAQVRRRRF 953
QY 1262 ERAIGPSQTHIRIQYS-PARLAYQDKGLVHNEVKVSLMRGLNPNVTSALSINIRKPD 1320
Db 954 ERAVGPFQTHIRIHYSVPMRVSYQDKGFLHNEVKVSLMRGLNPNVTSALSINIRKPD 1013
QY 1321 GYDYVAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 1014 GYDYVAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1054
RESULT 6
O77765 PRELIMINARY; PRT; 401 AA.
AC O77765;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Superficial zone protein (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Ruminantia; Bovidae.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99120896; PubMed=9920774;
RA Flannery C.R., Hughes C.E., Schumacher B.L., Tudor D., Aydelotte M.B.,
Kuettnner K.E., Caterson B.;
RT "Articular cartilage superficial zone protein (SZP) is homologous to
RT megakaryocyte stimulating factor precursor and is a multifunctional
RT proteoglycan with potential growth-promoting, cytoprotective, and
RT lubricating properties in cartilage metabolism.";
QY 671 PPTLKBPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPKGTAPTTTKEPAPTPPKPEP 730
Db 539 -----KEPEPTTPKKP-----EPTTKEPVPTPKPEP 565
QY 731 APPTPKGTAPTTLKEPAPTPPKAPKELAPTTTKEPTSTTSKDPAPTPPKETAPTPKPE 790
Db 566 EPTTP-----KEPEPTTPKKPEP-----TTRKEPEPTTPKEPEPTPKPE 604
QY 791 PAPTTPKKPAPTTPETPTTSEVSTPTTKEPTTHKGPDESTPLSAPETPKALENSP 850
Db 605 PEPTTPKKPEPTT-----617
QY 851 KPEQVPTTKTAAKPEMTTAKDKITERDLRTTBETTTAAPKWTKETATITTEKTTESKI 910
Db 618 -----TSPKT-----622
QY 911 TATTQTQVSTTTQDTPFKITLTKTTLAPKVVTTTKTITTTIMNKPSTETAKPKDRATN 970
Db 623 -----TTLKATTLAPKVTAPAE-----EIQNKPEETTPASEDSDD 657
QY 971 SKAT-----TPKP-QKPTKAPKKTSTKPKTPRVRKPKTTPRKRMTSTMPENLAP 1021
Db 658 SKTTLKPKQPKAPKPKTKPKKPTSTKPKT-PKTRKPKTTPAPLTKTSATPELNT 716
QY 1022 TSRIAEAMLQTTTRPNQTPNSKLVEVNPKSDAGAGETPHMLLRPHVFMPEVTPDMOY 1081
Db 717 TP--LEVMLPFTTIPKQPNPETAENVPDHEDADGEGEKP-LIPGPPVLFPPTAIPGTDL 773
QY 1082 LPRVFNQGIILNPLSDETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMFLSPSPSPARR 1141
Db 774 LAGRLNRGININPSPDETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMFLSPSPSPARR 833
QY 1142 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 1201
Db 834 ITEVWGIPSPIDVTRCNCEGKTFPFDKSDQYWRFTNDIKDAGYKPKIFKFGGLTGQIV 893
QY 1202 AALSTAKYKNWPSVYFFKRGSGIQYIYKQBEVQKCPGRRPALNYPVYGVEMTQVRRRRF 1261
Db 894 AALSIAQYKSRPSVYFFKRGSGVQYTYKQEPQCTQCTGRFPAINYSVYGETAQVRRRRF 953
QY 1262 ERAIGPSQTHIRIQYS-PARLAYQDKGLVHNEVKVSLMRGLNPNVTSALSINIRKPD 1320
Db 954 ERAVGPFQTHIRIHYSVPMRVSYQDKGFLHNEVKVSLMRGLNPNVTSALSINIRKPD 1013
QY 1321 GYDYVAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1361
Db 1014 GYDYVAFSKDQYINIDVPSRTARAITTRSGOTLSKWYNCP 1054

RL J. Biol. Chem. 269:2440-2446(1994).
 RN [2]
 RP SEQUENCE OF 626-1895 AND 4136-5179 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=93016075; PubMed=1400449;
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
 RA Kim Y.S.;
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
 RT both upstream and downstream of its central repetitive region.";
 RL J. Biol. Chem. 267:21375-21383(1992).
 RN [3]
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
 RX MEDLINE=93358717; PubMed=1885763;
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
 RA Petersen G.M., Kim Y.S.;
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
 RT and polymorphism";
 RL J. Clin. Invest. 88:1005-1013(1991).
 CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
 CC other mucus membrane-containing organs. Thought to provide a
 CC protective, lubricating barrier against particles and infectious
 CC agents at mucosal surfaces.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,
 CC bronchus, cervix and gall bladder.
 CC -!- PFM: All cysteine residues are involved in intrachain or
 CC interchain disulfide bonds (By similarity).
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 VWFC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions. There are no restrictions on its
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L21998; AB95295.1; -;
 CC EMBL; M74027; AA59875.1; -;
 CC EMBL; M94131; AA59163.1; -;
 CC EMBL; M94132; AA59164.1; -;
 CC PIR; A49963; A43932.
 CC Genew; HGNC:7512; MUC2.
 CC MIM; 158370;
 CC InterPro; IPR002919; Cysrich_TIL.
 CC InterPro; IPR006208; Cys_knot.
 CC InterPro; IPR006207; Cys_knot_C.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR001007; VWFC.
 CC InterPro; IPR001846; VWFC_D.
 CC Pfam; PF00007; Cys_knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; VWC; 1.
 CC Pfam; PF00094; VWD; 4.
 CC SMART; SM00041; CT; 1.
 CC SMART; SM00214; VWC; 2.
 CC SMART; SM00216; VWD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS00022; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01208; VWFC_1; 2.
 CC PROSITE; PS0184; VWFC_2; 2.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20 Potential.
 FT CHAIN 1 20 Mucin 2.
 FT DOMAIN 21 5179 Mucin 2.
 FT REPEAT 1401 1747 APPROXIMATE REPEATS.
 FT REPEAT 1401 1416 1.
 FT REPEAT 1417 1432 2.

FT REPEAT 1433 1448
 FT REPEAT 1449 1464
 FT REPEAT 1465 1471
 FT REPEAT 1472 1478
 FT REPEAT 1479 1494
 FT REPEAT 1495 1517
 FT REPEAT 1518 1533
 FT REPEAT 1534 1556
 FT REPEAT 1557 1572
 FT REPEAT 1573 1596
 FT REPEAT 1597 1612
 FT REPEAT 1613 1635
 FT REPEAT 1636 1651
 FT REPEAT 1652 1675
 FT REPEAT 1676 1693
 FT REPEAT 1694 1689
 FT REPEAT 1700 1715
 FT REPEAT 1716 1731
 FT REPEAT 1732 1747
 FT DOMAIN 4815 4886
 FT DOMAIN 4924 4991
 FT DOMAIN 5075 5160
 FT DISULFID 5075 5122
 FT DISULFID 5089 5136
 FT DISULFID 5098 5152
 FT DISULFID 5102 5154
 FT DISULFID ? 5159
 FT CARBOHYD 163 163
 FT CARBOHYD 423 423
 FT CARBOHYD 670 670
 FT CARBOHYD 770 770
 FT CARBOHYD 894 894
 FT CARBOHYD 1139 1139
 FT CARBOHYD 1154 1154
 FT CARBOHYD 1215 1215
 FT CARBOHYD 1230 1230
 FT CARBOHYD 1246 1246
 FT CARBOHYD 1787 1787
 FT CARBOHYD 1820 1820
 FT CARBOHYD 4339 4339
 FT CARBOHYD 4351 4351
 FT CARBOHYD 4362 4362
 FT CARBOHYD 4373 4373
 FT CARBOHYD 4422 4422
 FT CARBOHYD 4438 4438
 FT CARBOHYD 4502 4502
 FT CARBOHYD 4616 4616
 FT CARBOHYD 4627 4627
 FT CARBOHYD 4752 4752
 FT CARBOHYD 4787 4787
 FT CARBOHYD 4881 4881
 FT CARBOHYD 4888 4888
 FT CARBOHYD 4955 4955
 FT CARBOHYD 4970 4970
 FT CARBOHYD 5019 5019
 FT CARBOHYD 5038 5038
 FT CARBOHYD 5069 5069
 FT CONFLICT 1351 1351
 FT CONFLICT 1412 1412
 FT CONFLICT 1449 1449
 FT CONFLICT 1504 1504
 FT CONFLICT 4192 4192
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
 Query Match 19.3%; Score 1413.6; DB 1; Length 5179;
 Best Local Similarity 19.8%; Pred. No. 9.5e-20;
 Matches 461; Conservative 90; Mismatches 518; Indels 1254; Gaps 74;

QY 6 LPFYLLLSVFVIQVSSQDLSSCAGRCG-----EG-----YSR 40
 Db 502 LQVQLAPVWQLFVTLDDQASQ--GVQGLCGNFGNLEGGDDFKTAGSLVATGAGFANTWKA 559
 QY 41 DATCN-----CDY---48

```
Db 560 QSTCHDKLWLDDBPCLSNIESANYAEHWCSSLKKTETPFGRCHSAVDPAEYKRCCKDTC 619
QY 49 NCOHYMEC---CPDFKRVCTA-----ELSKC 71
Db 620 NCONNEDCLCAALSSVARACTAGVMLGMWREHVCKNVGSCFNSQVELYNLTTCQOTCR 679
QY 72 -----GRCFESF-----BRGR-----EADC-----DAQCKYDK 95
Db 680 SLSEADSHCLGAPFVPGGCPDHTFLDEKGRCVPLAKCSYHRGLYLEAGDVVVRQBER 739
QY 96 C-----96
Db 740 CVCRDGLHCRQIRLLIGQSTABKIHMDCSNLATATSKPRALSCQTLAAGYVHTCVSG 799
QY 97 --CPD-----YESFCAEVN-----109
Db 800 CVPDGLMDDRGCGVVEKECPCVHNNDLYSSGAKIKVDCTCTCKGRWVCTQAVCHGT 859
QY 110 -----109
Db 860 CSYSGHYITFDGKYYDFDGHCSYAVQYCGQNSLSGFSIITENVPCGTTGVTCSCA 919
QY 110 -----109
Db 920 IKIPMGRELKLEDKHRVVIQRDEGHVAYTTREVQYLVVSSSTGIIIVMDKETTVPFK 979
QY 110 -----PTSPSSKKAPPS- 123
Db 980 LAPYKGVTCGLCGNFDHRSNNDFTRDHMVVSSELDFGNSWKEAPTCPDVSSTNPEPCSL 1039
QY 124 -----GASQTIKST-----TKRSPK-----139
Db 1040 NPHRRWAEKQCSILKSSVFSICHKVDPKPFYACVHDSGSCDTGGDCFCFCASAVASYA 1099
QY 140 -----139
Db 1100 QECTKEGACVFWRPDLCPICFDYNNPPEHCEWHEPCGNRSFETCRTINGHSNISVSY 1159
QY 140 -----PNKKTKKVIIESEBEITEVDKNKNRTKK-----FTPKPP-----174
Db 1160 LEGCVPRCFKDRPIE-----EDLKKCVTADKCGCYVEDTHVPPGASVPEETCKSCV 1212
QY 175 -----VDEASGL-----DNG-----186
Db 1213 CTNSSQVVRPEEGKILNQDQGAFCYWEICGPNGTVEKHFNICSIITRPSTLITFTTIT 1272
QY 187 -----DEKVTTPDSTTQHNKVSTSPK-----208
Db 1273 LPTPTFTSTTTTTTPTSSVLSVTPKLCCLMSDWINEDHPSSGSDGDRPFGVCGA 1332
QY 209 -----208
Db 1333 PEDIECRSVKPHLSLEQHQKQVQCDVSVGFIKCKNEDQFNGPFGLCYDYKIRVNCWPM 1392
QY 209 -----ITTAKDINPRPSLPNSDTSKETSITVKEITVETKETTNNKQSTDCKEKTSK 265
Db 1393 DKCITTPSPPTTTPSPPTTTTTPPTTTPSPPT-----TTTTPPTTTPSPPTTTT 1447
QY 266 ETQSIKTSKADLAPTCKVLAKAPTKEATTTKGPAITPKPT-----PTTPKEPASITP 320
Db 1448 P-----LPTT-----TPSPISITTTTPPTTTPSPPTTTPSPPTTTPSPPTTT 1491
QY 321 KBPTPTTIKSAP--TTPKEP-----APTTPKSAP--TTPKEPAPTTPKEPAPTTPKEP 369
Db 1492 TTPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1551
QY 370 -----APTTPKEPAPTTPKSAP--TTPKEPAPTTPKAPAPTTPKEPAPTTPKEP 422
Db 1552 STTTLPTTTPSPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTP 1611
QY 423 EPAPTTPKEPAPTTPKEP-----APTAPKAPAPTTPKEPAPTTPKEPAPTTPKEPSP 475
```

```
Db 1612 PPPTTTPSPPTTTPPTTSTTLTTPPTTTPSPSEPPTTTTTPPTTTPSPPTTTPSPPTT 1671
QY 476 PKEPAPTTPKSAPTTTPKEPAPTTPKSAP---TTPKEPSPPTTTPKEPAPTTPKEPAPTTPKK 532
Db 1672 TTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSSPI-TTTPSSSITTPSPPTTTPMT 1730
QY 533 PAPTTPKEPAPTTPKEPAPTTPKAPAPTTPKAPAPTTP-----KETAPTTPKKLTP-----583
Db 1731 PSPTTTPSPPTTTPMTLPLPTTSSPLTTPPLPSITPTTPSPPTTTPPTTP--CVPLCNW 1788
QY 584 -----583
Db 1789 TGWLDSGKFNHFKPGGDELIGDVCGPGWAANISCRATMYDPVPIGOLGQTVVCDVSVGL 1848
QY 584 -----TTEKLAAPTTPPKAPAPTTPBEELAPTTPPEE 612
Db 1849 ICKNEQKPGGVIPIMAFCLNVEINVOCCECVQPTMTTTTTTENPTPTTPTTPTTPTT 1908
QY 613 PTP-----TTPPEPAPT-----TPKAAAPT-----634
Db 1909 PTPPTGQTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1968
QY 635 -----KEPAPTTPKEP-----APTTPKEPAPT--TPKETAPTTPKAPTTPKAPAPT 679
Db 1969 PTTTTTPTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2028
QY 680 TTPKAPKAPKELAPTTPKEP-----TSTSDKAPAPT-----709
Db 2029 TGTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2088
QY 710 -----TPKGT-----APTTPKEPAPT-----TTPKEPAPT-----733
Db 2089 TTVPTPTPTGQTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2148
QY 734 -----TPKGT--APT--LKEPAPTTPKPKAPKELAPTTPKGP 767
Db 2149 PTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2208
QY 768 -----TSTSDKAPAPT-----TTPKETAPTTPKEPAPT-----793
Db 2209 TPTPTGQTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2268
QY 794 ---TTPKAPAPT---TPETPP-----PTTSEVS-----TPT 818
Db 2269 ITTTTTTPTPTPTGQTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2328
QY 819 TTKPTTTHKSPDSETPSLSAEPTPKALENSPKEP-----GVPT-TKPTAATKPE 867
Db 2329 GTGQPTT---PTTTTTTPTPTPTGQTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2385
QY 868 MTITAKDKTTERDLR---TTP---ETTTAAPKMT---KETATTEKTTSKSI-----910
Db 2386 TTTTPTPTPTGQTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2445
QY 911 -TATTTQVSTT-----TQDTPPKITTT-----LKTTL 938
Db 2446 QPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2505
QY 939 APKVTTT-KKTLITTEIMNKEBEETAKPKDRATNSKATTP-----XPKQPK 983
Db 2506 TPTPTGQTGPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2565
QY 984 AP-----KKPTSTKKKXTMPVRVKPTTPTTPKMTSMPELNPSTRAEAMLOTT 1033
Db 2566 TPTTTTTTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2617
QY 1034 T-----RNOPTNSKLVEVNKSEADAG-----GARG 1059
Db 2618 TTVPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2677
QY 1060 ETPHMLLRPHVFWPEVTPDMDYLPRVNOGIIINPMLSDETNICNGKPVGDLTTLANGTL 1119
Db 2678 PTTTPTTTTTTPTPTPTGQTPTT-----TPTTTTTT-----2711
```

QY 1120 VAFRGHYFWMUSPFFSP -----SPARRITEVWGIESPIDT 1154

Db 2712 -----VPRTPTGTGQTPTTTTPITTTTTPPTPTGT 2744

RESULT 8

Q9VR49 PRELIMINARY; PRT; 1225 AA.

ID Q9VR49

AC Q9VR49

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)

DE CG3047-PA.

GN Name=SGS1; ORFNames=CG3047;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prerygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beson K.Y., Benos P.V., Berman B.P., Bhargava M., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foeiser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M.K., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

Yeh R.F., Zaveri J.S., Zhao M., Zhang G., Zhou Q., Zheng L.,

Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,

Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of Drosophila melanogaster.";

Science 287:2185-2195(2000).

[2]

RN SEQUENCE FROM N.A.

RX MEDLINE=22426065; PubMed=12537568;

RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,

RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,

RA George R.A., Hoskins R.A., Laverty T., Munzy D.M., Nelson C.R.,

RA Pacleeb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,

RA Svirskaas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,

Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;

"Finishing a whole-genome shotgun: release 3 of the Drosophila

melanogaster euchromatic genome sequence.";

Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

RL

| | | | |
|--|--|---|-----------------------------------|
| QY | 1122 | FRGHVFWLSPSPSPARRITEVWGIPSPIDVTFRNCBCKTFFFKDSQVWRFTNDIK | 1181 |
| Db | 1698 | -----PLPLPLP-----NP-PLPLPEI----- | 1703 |
| QY | 1182 | DAGYKPKIPKGGGLTGQVIALSTAKYNWPSVYFFKRGSIQYIYKQBPVKCPGR | 1241 |
| Db | 1704 | -----PLPL-----PPLPTA-----LPLPLPLPL | 1723 |
| QY | 1242 | REALNVPVYGVMTQVRRRFRFAIGBSQTHIRIQYSPARLAYQDKGLHNEVKVLSILWR | 1301 |
| Db | 1724 | PP----- | 1725 |
| QY | 1302 | GLFNVVTSAISLPNIRKPD | 1320 |
| Db | 1726 | LPEVNLTAISLPEISLPN | 1743 |
| RESULT 10 | | | |
| AAS64673 | | | |
| ID | AAS64673 | PRELIMINARY; | PRT; 1761 AA. |
| AC | AAS64673; | | |
| DT | 01-APR-2004 | (TrEMBLrel. 27, Created) | |
| DT | 01-APR-2004 | (TrEMBLrel. 27, Last sequence update) | |
| DT | 01-APR-2004 | (TrEMBLrel. 27, Last annotation update) | |
| DE | CG33300-PA | (Fragment) | |
| GN | CG33300. | | |
| OS | Drosophila melanogaster | (Fruit fly). | |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | |
| OC | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | |
| OX | NCBI_TaxID=7227; | | |
| [1] | | | |
| SEQUENCE FROM N.A. | | | |
| RP | MEDLINE=20196006; | PubMed=10731132; | |
| RX | Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., | | |
| RA | Anatitides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F., | | |
| RA | George R.A., Lewis S.J., Richards S., Ashburner M., Henderson S.N., | | |
| RA | Sutton G.G., Wortman J.R., Yang S.P., Zhang Q., Chen L.X., D., | | |
| RA | Brandon R.G., Rogers Y.H., Blazer R.G., Champagne M., Pfeiffer B.D., | | |
| RA | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., | | |
| RA | Abriel J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D., | | |
| RA | Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | | |
| RA | Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., | | |
| RA | Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P., | | |
| RA | Burtis K.C., Buesan D.A., Butler H., Cadieu E., Center A., Chandra I., | | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | | |
| RA | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., | | |
| RA | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | | |
| RA | Durbin K.J., Evangelista C.C., Ferraz C., Fierera S., Fleischmann W., | | |
| RA | Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., | | |
| RA | Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., | | |
| RA | Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., | | |
| RA | Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., | | |
| RA | Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., | | |
| RA | Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X., | | |
| RA | Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., | | |
| RA | Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., | | |
| RA | Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., | | |
| RA | Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., | | |
| RA | Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., | | |
| RA | Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., | | |
| RA | Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., | | |
| RA | Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.F., Smith T., | | |
| RA | Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., | | |
| RA | Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., | | |
| RA | Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J., | | |
| RA | Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., | | |
| RA | Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.H.O., | | |
| RA | Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., | | |
| RA | Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; | | |
| RT | "The genome sequence of Drosophila melanogaster." | | |
| RL | Science 287:2165-2195(2000). | | |
| [2] | | | |
| SEQUENCE FROM N.A. | | | |
| RP | MEDLINE=22426065; | PubMed=12537568; | |
| RX | Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A., | | |
| RA | Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A., | | |
| RA | George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R., | | |
| RA | Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J., | | |
| RA | Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C., | | |
| RA | Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.; | | |
| RT | "Finishing a whole-genome shotgun: release 3 of the Drosophila | | |
| RT | melanogaster euchromatic genome sequence." | | |
| RL | Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002). | | |
| [3] | | | |
| SEQUENCE FROM N.A. | | | |
| RP | MEDLINE=22426069; | PubMed=12537572; | |
| RX | Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., | | |
| RA | Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E., | | |
| RA | Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P., | | |
| RA | Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A., | | |
| RA | Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., | | |
| RA | Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., | | |
| RA | Lewis S.E.; | | |
| RT | "Annotation of the Drosophila melanogaster euchromatic genome: a | | |
| RT | systematic review." | | |
| RL | Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002). | | |
| [4] | | | |
| SEQUENCE FROM N.A. | | | |
| RP | MEDLINE=22426070; | PubMed=12537573; | |
| RX | Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R., | | |
| RA | Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M., | | |
| RA | Ashburner M., Celniker S.E.; | | |
| RT | "The transposable elements of the Drosophila melanogaster euchromatin: | | |
| RT | a genomics perspective." | | |
| RL | Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002). | | |
| [5] | | | |
| SEQUENCE FROM N.A. | | | |
| RP | FlyBase; | | |
| RL | Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases. | | |
| [6] | | | |
| SEQUENCE FROM N.A. | | | |
| RP | FlyBase; | | |
| RA | Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; AE003626; AAS64673.1; - | | |
| FT | NON TER 1 | | |
| SQ | SEQUENCE 1761 AA; 197521 MW; 62A5E16E6241B3F4 CRC64; | | |
| Query Match 18.3%; Score 1339.8; DB 2; Length 1761; | | | |
| Best Local Similarity 24.1%; Pred. No. 4.1e-19; | | | |
| Matches 482; Conservative 106; Mismatches 399; Indels 1012; Gaps 94; | | | |
| QY | 74 | CF----- | ESFERGECDCDAQCKYKCCPDYSEFCAEV 107 |
| Db | 5 | CFIWLALVHSANAALGSEIKFGDESTTESAE-----KY-----PDYCW----- 44 | |
| QY | 108 | HNPTSPSSKKAPPPSGASQIKTKTSKSPFPNNKKTKKVISEPE-----ITVKDNKK 162 | |
| Db | 45 | NPELPGC-----PFGEGDGNNTTKISTLATTK-----STVTSEETTLTKITIKSTAK 94 | |
| QY | 163 | NRTKK-----KPTPKPVPVDEAGSLDNGFKVTPDTST-----TOHNKV----- 203 | |
| Db | 95 | PTTKTNTNEPTE-----KITTAKATIKSTATTARATAPKTEQTTLR 138 | |
| QY | 204 | -----STSPKITTAKPINPRLPNSDTSKET-----SLTVNKKETVTETKTTTINK-OTSTD 256 | |
| Db | 139 | TTTIKSTSELITTLK-----TTTIKSTAKTSKSTHNTTKSTLTATTEPTTRKSTAKT 194 | |
| QY | 257 | GHEKTTSAKETQ-----SIEKTSKADLAPTSKVLAKETPKPAFTTKGPAFTTKPEPT-- 308 | |
| Db | 195 | TREPTTKKETTERTTQBPSTSKTTTHTHE--TTAEPATKKTTHPTTKQKSTTLRIEPTTR 252 | |
| QY | 309 | -----PTTPKEPASTTPKEPT-----PTTIKSA----- 331 | |
| Db | 253 | KSSTAKTTPETTKETTERTTKSTKTTTHTTETTABEPATKKTTHPTTKQKSTTLRIE 312 | |

Wed Oct 13 12:38:27 2004

1225 IQOYIKQEPVQKCPRRPALNYPVYGMTQVRRRFFERAIGPSQTHIRIQSPARLAY 1284
 121 IQOYIKQEPVQKCPRRPALNYPVYGMTQVRRRFFERAIGPSQTHIRIQSPARLAY 180
 1285 QDKGVLHNEVKVSIWRLGPNVVTSAISLNPTRKPGYDYAFSKDOYNNIDVPSRTARA 1344
 181 QDKGVLHNEVKVSIWRLGPNVVTSAISLNPTRKPGYDYAFSKDOYNNIDVPSRTARA 240
 1345 IITRSQOTLSK 1355
 241 IITRSQOTLKF 251

RESULT 12
 Q7PMD5 PRELIMINARY; PRT; 3150 AA.
 AC Q7PMD5;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE ENSANGP0000004555 (Fragment).
 OS Name=ENSANGP00000003651; PEST.
 GN Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AA0801008980; EAL13969.2; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
 DR InterPro; IPR006770; OGF_recept.
 DR Pfam; PF04680; OGFr_III; 80.
 FT NON TER 1
 FT NON TER 3150
 SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441CB8C839 CRC64;
 Query Match 17.9%; Score 1310.1; DB 2; Length 3150;
 Best Local Similarity 19.1%; Pred. No. 5.1e-18;
 Matches 465; Conservative 130; Mismatches 490; Indels 1349; Gaps 96;
 QY 110 PTSPP-----SSKKAPPPSGASQTIKSTTKSPKPPNKKTKKIESEIEVKNKN 163
 DB 141 PRPTPTDTMSSASTPEPS-----TTPGTRTTRTPTPTPTDTMSSASTPEPSPTTPTGTR 196
 QY 164 RTKKKPTP-----KPVVDEAGSLDNGDFKVTTPOTST 198
 DB 197 TTPTRPTPTDTMSSASTPEPSPTTPTGTRTTRTPTPTPTDSTMS-----SASTPEPSTT 249
 QY 199 QHN-----KVESTPKITAKPINP----- 217
 DB 250 PGTRTTRPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPE 309
 QY 218 -----RPLSPNSQTSKETSLSVNKEVVEIKETTKNQSTSD----- 256
 DB 310 PSTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSAST 368
 QY 257 -----GKEKTSK--EQOSTEKSADLAP-----TSKVLAKPTPKAET----- 294
 DB 369 PEPSTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSS 428
 QY 295 -----TTKGPALATPKPEPTP-----TTKPEPAS----- 317
 DB 429 SASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTM 488

318 -----TTPKEP-----TPTTKSAPTTP----- 335
 DB 489 SSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTEST 548
 QY 336 -----KEPA--PTTKSAPTTP-----KEPAPTTPKEPAPTTPKEPAP 371
 DB 549 DTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPT 608
 QY 372 T-----TTKEPA--PTTKSAPTTPKEPAPTTPKKKAPT-----TTPKEP 408
 DB 609 TESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPT 667
 QY 409 AP-----TTPKEPTP-----TTPKEAPT-----TKE 430
 DB 668 SITPTGTRTTRTPTPTDSTMSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPE 727
 QY 431 PAP-----TTPKEAPT-----APKKAPTTPKEPAPTTPKEPAPT-----TTK 469
 DB 728 PSTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSAST 787
 QY 470 EPSPT-----TTPKEPA-----PTTKSAPTTP-----PTTKSAPTTP----- 490
 DB 788 EPSPTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSS 847
 QY 491 --TKEPA--PTTKSAPTTP-----KEPSPTTPKEPAPTTPKEPAP----- 527
 DB 848 ASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTEST 907
 QY 528 -----TTPKKAPAP-----TTPKEPAP----- 543
 DB 908 SASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTEST 967
 QY 544 -----TTPKEPA-----PTTKSAPTTP-----PTTKSAPTTP----- 554
 DB 968 TTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPT 1027
 QY 555 K-----KEPAPT-----KEPAPTTPKETAPTTPKK 580
 DB 1028 ESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPT 1087
 QY 581 LTPT-----TPE-----KLAPT-----TPE----- 595
 DB 1088 PTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRT 1147
 QY 596 -----KEPAPTPELAPTTPPEPTPT-----TPE----- 619
 DB 1148 TPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTG 1207
 QY 620 -----EPAPTTPKAAAPTTPKEPAP----- 639
 DB 1208 TTRTPTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPT 1267
 QY 640 -----TTPKEP-----APTTP-----KEPAP 655
 DB 1268 GTTRTPTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPS 1327
 QY 656 TTPKETAPTTPKGTAPTTLK-----PAP-----TTPKKPAPKELAPT----- 694
 DB 1328 TTPPTDTRTTP--TRPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSS 1385
 QY 695 --TKPEPTST-----TSKAPAPTTPKGTAPTTPKEP----- 722
 DB 1386 ASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTESTDT 1445
 QY 723 -----APTTPKE-----PAPTTPKGTAPTTLKAPAP-- 748
 DB 1446 TMSASTPEPSPTTPTGTRTTRTPTPTSTESTDTMSSASTPEPSPTTPTGTRTTRTPTPTSTEST 1505
 QY 749 -----TTPKKPAPKELA----- 760
 DB 1506 STDTMSSAVTPDPSGIPGTRTTRTPTPTPTETMSSASTPEPSMTPTGTRTTRTPTPTST 1565
 QY 761 -----PTTKGPTSTTSDK 774

```
Db 1566 ESTDTMSSASTPEPSTTGGTTRTTTTPRPTSTESTDTTMSSASTPEPSTTGGTTRTTPR 1625
QY 775 PA-----PTTPKETAPTTPKPAP-----TTPK 797
Db 1626 PTSTESTDTMSSASTPEPSTTGGTTRTTTTPRPTDTTMSSASTPEPSTTGGTTRTTPR 1685
QY 798 KAPT-----TPE--TTPPTT-----SEVSTPTT 820
Db 1686 RPTPTDSTMSSMSSESTPEPSTTTPSTTGGTTRTTTTPRPTSTESTDTTMSSASTPEP 1745
QY 821 KPTTIHKSP-----DESPPELSABP-----841
Db 1746 TTPGTRTTPRPTSTESTDTTMSSASTPEPSTTGGTTRTTTTPRPTSTESTDTTMSSAST 1805
QY 842 ----TPKALENSKPEP---GVPTTKPAANKPEMT-----870
Db 1806 PEPSTTGGTTRTTPRPTSTESTDTTMSSASTPEPSTTGGTTRTTTTPRPTSTESTDT 1865
QY 871 ---TAKDKTTERDLATP-----ETTTAAPKMTKETA 899
Db 1866 SASTPEPSTTGGTTRTTTTPRPTSTESTDTTMSSASTPEPSTTGGTTRTTTTPRPTSTESTD 1925
QY 900 TT-----TEKTESKITATTQVSTTTQD-----TTPFKIT 931
Db 1926 TTMSSASTPDSTTGGTTRTTTTPRPTSTESTDTTMSSASTPEPSTTGGTTRTTTTPRPTSTEST 1985
QY 932 ----TLKTTTLAPKVTTKTKTITTEIMNKPEETAKP-----964
Db 1986 DTMSSASTPEPSTTGGTTRTTPRP--TSTERDTTMSSAYTPEPSTTGGTTRTTPR 2043
QY 965 ----KDRATNSKATTPKQO-----KPTK-----APKK 987
Db 2044 TSTESTDTTMSSASTPEPSTTGGTTRTTPRPTSTESTDTTMSSASTPEPSTTGGTTRTTPR 2103
QY 988 PTSTK-----KPKTMPVRK-----PKTTPTRK 1011
Db 2104 PTSTESTDTTMSSASTPEPSTTGGTTRTTTTPRPTSTESTDTTMSSASTPEPSTTGGTTR 2163
QY 1012 MTSTMPELNPT-GRIAEALQTTTRPQNTNSKLVEVNPKSEDAAGAGETPHMLLRPHV 1070
Db 2164 TTTPRP--TPTDSTMSSMSASTTGGTTRTTTTPRPTSTESTDTTMSSASTPE-----2214
QY 1071 FMEVTPDMOYLPRVFNQGHIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWM 1130
Db 2215 --PSTTGP-----TTRTTPRPTPD--STMSS-----M 2240
QY 1131 SPSPSPSPA-----RRI TEVWGI PSLDITVTRCNCEGKTFPFKDSQYWRFTNDIKDAGY 1185
Db 2241 SSASTPEPSTTGGTTRTTPR--RPTPTDIT--MSSSMSSAST 2278
QY 1186 PKPIFKGFGGLT-----GQIVAAALSTAKYKNWPSVYFFKGGSIQO-----1227
Db 2279 PEP--STTPTGTWTTTTPRPTSTESTDTTMSSMSASTPESTTPE-----LLQDDOHL 2327
QY 1228 ----YIYKOE-----FVQKCPGRRPALNVPVVGEMTQVRRR 1260
Db 2328 PIPQCLVKCRQPLLRGHRHLVQPELLQDDOHLPIQC--RQPLL-----RSH 2373
QY 1261 FERAIGPSQTHIRIQYSPARLAYQKGVULHNEKVSVILWRGLPNVVT--SAISLPINRK 1318
Db 2374 LRHLVQF-----ELLQDD-----QHLPSDTTMSSASTPE---2403
QY 1319 PDGYDYAFSKQYVNIIDVPSRTARAITRSQT 1352
Db 2404 -----PSTTPG--TTRTTP 2416
```

RESULT 13

Q6SSE6

ID Q6SSE6 PRELIMINARY; PRT; 3409 AA.

AC Q6SSE6;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

```
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
GN Plus agglutinin.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_Taxid=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.; 2003 to the EMBL/GenBank/DBJ databases.
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AAS07044.1; -.
DR InterPro; IPR002951; Atrophin.
DR PRINTS; PR01222; ATROPHIN_extensin.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 17.2%; Score 1256.8; DB 2; Length 3409;
Best Local Similarity 14.0%; Pred. No. 7.5e-17;
Matches 398; Conservative 166; Mismatches 460; Indels 1812; Gaps 80;

QY 1 MAWKTLPI-----YLLLLLSV-----16
Db 50 LAWVTIPVDALQFFVNVSVTSSAATDAPSAIYIGALLDLVLMVWPQNCINDTRRSYD 109
QY 17 -----16
Db 110 APWFSRCALPAVQYGYDEYDIDATVYSGNSLRPFYSTCRYPASSDPYAFLEPYSMWVD 169
QY 17 -----FVIOQVSS-----OOLSSACRCGEGYRD 41
Db 170 DFGIYQVLDGDFMWWGFTVNTVDSNLKWDPAWTAQAMLGQDF-----RD 217
QY 42 ATCN-----CDYNCQHVMCECPDFKRVCTAELSCCKCFESFERGECDCDAOCKKYD-- 94
Db 218 AVNMFQHYCSW--PFVRC-----SSCENIDIA 243
QY 95 -----94
Db 244 DPYDPDKIANGIYPAVITDALDFRNASLIYIYLDFGMFPAGSLDNVWYINLAYNFIGGLP 303
QY 95 -----KCCPD-----YE-----101
Db 304 ANLPTLLPSLQHLALDHCRATPDVGTASLOYQWQVYEGSDGVTGSDGTEY 363
QY 102 -----101
Db 364 VISGMIPEWGDVAAASDSSLPLWANLRTVRLSNQALYGIPEGLRSASSISSWELQ 423
QY 102 -----SFCAEVHN-----109
Db 424 NTELOGLPEFAAPINSLLYLGKLTHTWVMDPADHTQGGECLAPPSPSPSPRPR 483
QY 110 -----PTSPSSKKAP-----PPSGASOT-----128
Db 484 PEPPLPSPPPPLLPSPVPVPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 543
QY 129 -----IKSTTKRSPKPNKKTKKVIIESEITEVKONKNTKKKPTP 171
Db 544 GICDSFSPMPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 578
QY 172 KPPVVDVDEAGSLDNGDFKVTTPDT---STQHNKYSTSPKITTAKPINRPSLPNSDTS 228
Db 579 SPP-----FHPSPDPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 620
QY 229 KETSLTVNKETTETKTTTINKQISTDGKENTTSKETQSIETKSADLAPTSLKAP 288
Db 621 -----PPSPAPSP 629
QY 289 TPKAETTTKGPALTPPKETPTTPKEPAST-----TPKEPTTIKSAPTTPKEPAPT 341
```

| | | | | | | | |
|----|------|--|------|----|------|--|------|
| Db | 630 | APPS-----PAPSPSPSPVPQPPSPVPPSPKPPSPAPPSPVPSPAPPSPAPPSPAPSPSA | 683 | Db | 1694 | SGSSSLSGAATANOQQRHVHFLLPATAYADAARNPGRNDLSLSVELTDNAVASPAVGEA | 1753 |
| Qy | 342 | TTKSAPTTPKEPAPTTPKAPTTPK-----EPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTP | 397 | Qy | 912 | -ATTQVTTTITQDTPFPKITTTLKTTTLAPKVT----- | 943 |
| Db | 684 | PWPAPSPAPPLPSPSPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP | 743 | Db | 1754 | LATTARVTAATYP-----AVAATTLVAASSSSFAQAIKRAKSGLLQSGSYHIQMLTMSL | 1806 |
| Qy | 398 | KKPAPT-----TPKEPAPTTPKEPTTPTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT | 442 | Qy | 944 | -----TTKKTITTTT----- | 953 |
| Db | 744 | EPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP | 803 | Db | 1807 | YLASRGVGREYGRYAVEFKYAVLGVKGNLGPABEAAMPITNEKEVTAAEQARQVGGDLWPIG | 1866 |
| Qy | 443 | ---APKKPAPTTPKEPAPTTPKEPAPT-----ITKE | 470 | Qy | 954 | ---IMNKPETAKPKORATNSKATTPKQKP----- | 981 |
| Db | 804 | PSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP | 863 | Db | 1867 | NDLGGSSNTTASGSSSGSSSSNSP-PRRPPPPPAAGSTGLLFSNADASPPLAVATPA | 1925 |
| Qy | 471 | PSPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT | 527 | Qy | 982 | -----TKAPKPTSTKKPKTPRVRKPKTTPTPRKMST--MPELN----- | 1020 |
| Db | 864 | PAPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP | 922 | Db | 1926 | LPAPLSTTIAAATAAP-----PRLPSP--PPPAVGSSTGVLPRRHLMQOML | 1970 |
| Qy | 528 | -----TPPKPAPTTPKEPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK | 563 | Qy | 1021 | -----PTSRIAEAMLOTTTRPNQTPNSKLVE----- | 1046 |
| Db | 923 | SPEPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP | 982 | Db | 1971 | QPPAAAVAAPPPPPFASSSALVLPSPPPPPPPSQLLIQQASATVYVSDMDLLYTLVVAA | 2030 |
| Qy | 564 | -----EPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK | 608 | Qy | 1047 | -----VNPKSEDAGGAEGETPHMLLRPHVFMPEVTPMDYLPVFN | 1087 |
| Db | 983 | PPGAPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP | 1041 | Db | 2031 | LITAVAAGRLIAAVLYRLVLSPE-----PHPELA-----FPRLE-2064 | |
| Qy | 609 | -----TPPEPTTPPEEPAPTTPKAAAPN-----TPKEPAPTTPKEPAPT | 648 | Qy | 1088 | QGIINEMLSDETNICNGKPVDTGLTLRNGTLVAFRGH----- | 1125 |
| Db | 1042 | PP | 1101 | Db | 2065 | -----TTIAGLILVALTFYSCMALGPAADWEGSRSTAAYCV | 2100 |
| Qy | 649 | TPK-----EPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP | 674 | Qy | 1126 | -----YFWMLSPFPSPARRITEVWGIP-----SPIDTVFTRCN-1160 | |
| Db | 1102 | PPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP | 1161 | Db | 2101 | LTIADVVAFLMWLA-----LARAMVVPQFTLVEPMTTSSVASPRPSAFERADR | 2150 |
| Qy | 675 | -----KEPAPTTPKAPKELAPT-----TKPTSTITS | 704 | Qy | 1161 | -----CEGKTFFFKDSQY----- | 1173 |
| Db | 1162 | PPTPP | 1221 | Db | 2151 | TSRPSVTSIATAADIGTATGGGGGAASGACADVGAAPATVAAGADVSEDDRYARG | 2210 |
| Qy | 705 | KPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPT | 763 | Qy | 1174 | --WR-FTNDIKAGYKPIFKFGGLTGQIIVAALSTAKYKNPESVYFFKRGSGSIQYIY | 1230 |
| Db | 1222 | SPAPPSPAPPSP-APPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP | 1276 | Db | 2211 | PHWKQF-----DGLVPTTVASGVGSGGVVPL----- | 2238 |
| Qy | 764 | TKGPTSTSDKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK | 822 | Qy | 1231 | KOEPVQKCPGRRPALNYP-----VYGEWTVQVRRRFPERRA | 1264 |
| Db | 1277 | SPAPPSPAPPSPAPPSP---APPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP | 1333 | Db | 2239 | --PELVALPGSGRHLPLPPLPAAAGATAAGGAATAPGSPRGAAGE----- | 2284 |
| Qy | 823 | PTTIHKSPESTBELSABPTPKALE-----NSPKPBGV---PTT | 858 | Qy | 1265 | IGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWR----- | 1301 |
| Db | 1334 | PLPSPAPPPLVPVPASPAPPSPPLPPQOTPAMPPPPPPPPPPPPPPPPPPPPPPPPP | 1393 | Db | 2285 | ---DTH---QFGP-----H-----WRKESVPDDGAATAGVGDAAAGGA | 2317 |
| Qy | 859 | KTP-----AATKPE----- | 867 | Qy | 1302 | ----- | 1301 |
| Db | 1394 | PTPPLAPLPDCTLLAQALLSI PDAANSVFVVSAGLPTS VAPSTPEPELLASFCVCSQ | 1453 | Db | 2318 | GGGGRGSRGSGTDGVRFANGTRTPTSDGAKGRSSSHGGGNGNAMSGAASFGASAGG | 2377 |
| Qy | 868 | MTTAA----- | 872 | Qy | 1302 | -----GLP-----NVVTSAISL----- | 1313 |
| Db | 1454 | LTATAISLVGSSRGNTNGSSGGNYVNGGDAAIQRIAGNSTDRGSSSGSGSSSW | 1513 | Db | 2378 | GAASASRLGSRSGSPGDGDYNAMSSGAAMRGGAACASRPGPAGSRAGSPPEGAGGVAGGA | 2437 |
| Qy | 873 | -----KDKTTERDL-----RTTP----- | 885 | Qy | 1314 | ----- | 1319 |
| Db | 1514 | GPGETAAEWAVDAVDQGYIQLSIGGYTYTTVVVDRTTPSVSGNVTLNRIKQEPS | 1573 | Db | 2438 | NAMFAGTAGAGAISPAHDGSRSPGSGNEGSRQMSYGNAMTAGAVGVGGVRRPAG | 2497 |
| Qy | 886 | -----ETTTAAPKVT----- | 895 | Qy | 1320 | -----DGYDYVAFSKDQYNTIDVPS-----RTARAI----- | 1345 |
| Db | 1574 | AVGEASLNALGSKQAMLLTISTEPVPAFDPAASLIVTGALVAEWAAADKMTFYVIAMT | 1633 | Db | 2498 | SLENASDG-----DAPPLPAPGTVPPLPFANLLTSGADVESRRLNAEAVA | 2543 |
| Qy | 896 | ---KETATTEKTESKIT----- | 911 | Qy | 1346 | -----TTRSG 1350 | |
| Db | 1634 | LPAELVATAAGSSSSSGTSRSGNGNGTAAAAAAPPAPPGITGRRRALQOQAAAPPFA | 1693 | Db | 2544 | AAAAAGGGGGGTSRSG 2559 | |
| Qy | 912 | ----- | 911 | | | | |

```

ID AAS07044 PRELIMINARY; PRT; 3409 AA.
AC AAS07044;
DT 02-MAR-2004 (TEMBLrel. 27, Created)
DT 02-MAR-2004 (TEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TEMBLrel. 27, Last annotation update)
DE Plus agglutinin.
GN SAG1.
GE Chlamydomonas reinhardtii.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]_
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RT "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV45930; AAS07044.1;
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 17.2%; Score 1256.8; DB 2; Length 3409;
Best Local Similarity 14.0%; Pred. No. 7.5e-17;
Matches 398; Conservative 166; Mismatches 460; Indels 1812; Gaps 80;

QY 1 MAWKTLPI-----YLLLLLSV----- 16
Db 50 LAWVTIPVDALQFVNVSVTSSAATDAPSAIAYIGALLDVLWNVPWQPNCLNDRRSYD 109
QY 17 ----- 16
Db 110 APWSRCALPAVQGYDEYDIDATVYSGNSLRPFYSTCRYPASSDPAFLPYMPWVD 169
QY 17 -----FVIQVSS-----QDLSSCAGCGEYRD 41
Db 170 DGIYQPVLDGDMWGFVNTVDSNLKWDIPANTAQAMLGQDF-----RD 217
QY 42 ATCN-----CDYNQHWECPCDFKRVCTAELSCKRCFESFERGECDCDAQCKYD-- 94
Db 218 AVNFGTHYCSW---PFVEC-----SSCENVDIA 243
QY 95 ----- 94
Db 244 DPYDPLKIANGIVPAVITALDFRNASLYIYLDGFMFPAGSLDNVWYINLAYNFIGGLP 303
QY 95 -----KCCPD-----YE----- 101
Db 304 ANLPTLLPSLQHLALDHCRAITDVRGTASLOYGWQVPTGQPYEYCSGDGVTGSDGTEY 363
QY 102 ----- 101
Db 364 VISGMIPDEWGDVAAAASDSLPLWANLRTVRLSNQALYGPPEGLRSASSISSWRLQ 423
QY 102 -----SFCAEVHN----- 109
Db 424 NTELCGLPEFAAPINSLLYLTKLGHVTHWMDPADHTQGECLAPPFPSPSPRPPR 483
QY 110 -----PTSPSSKXAP-----PPSGASOT----- 128
Db 484 PPPLPPSPPLPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 543
QY 129 -----IKSTTKRSPKPNKKTKKVISEIEITEVDKNKNTKKKPTP 171
Db 544 GICDPSFPMPPSPRPPSPRPPSPRPPSPRPPSPRPPSPRPPSPRPPSPRPPSPRPPSP 578
QY 172 KPWVDEAGSLDNGFKVTTPTD---STQHNKVSSTPKITTAQINPRPSLPNSDTS 228
Db 579 SPP-----FHPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 620
QY 229 KETSLTVNKETTVEKTTTNNKQTDGDKETTSAKETQSIKTSKOLAPTSKVLAP 288
Db 621 -----PPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 629
QY 289 TPKAETTTKGPALTTKPEFTTPKEPAST-----TPKEFTTTKAPTTPKEPAPT 341

```

```

Db 630 APPS-----PAPSPQSPSPVPVPPQPPSPVPSPKPPSPAPSPVPSPSPAPPSPAPSPA 683
QY 342 TTKSAFTTKKEPAPTTTKKEPAPTTTK-----EAPITTKKEPAPTTTKSAFTTKKEPAPTTT 397
Db 684 PNPAPPPAPPLPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 743
QY 398 KKPAPT-----TPKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPT-- 442
Db 744 EPPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPAP 803
QY 443 ---APKKAPTTKKEPAPTTTKKEPAPT-----TTKE 470
Db 804 PNPAPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 863
QY 471 PPTTKKEPAPTTTKSAFTTKKEPAPTTTKSAFTTKKEPAPTTTKSAFTTKKEPAPTTTK 527
Db 864 PAPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 922
QY 528 -----TTPKAPAPTTPKEPAPTTTKKEPAPTTTKKAPAPTAK----- 563
Db 923 SPEPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 982
QY 564 -----EAPITTKETAPTTPKKLTPTTPEKLAFTPE--KPAPTTPPELAPT- 608
Db 983 PPSAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1041
QY 609 -----TPEPTTTPKEPAPTTTKKAAAPN-----TPKEPAPTTTKKEPAPT 648
Db 1042 PPPPPPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1101
QY 649 TPK-----EPAPTTPKETAPTTPKGTAPTTL-----TKETSTTSD 704
Db 1102 PPSPEPPSPAPSPSPSPSPSPSPSPSPSPSPAPLLPPSPDPSPSPAPSPMPPLPTSPSPSP 1161
QY 675 -----KEPATTTPKKAPKELAPTT----- 1221
Db 1162 PPTPPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1221
QY 705 KPAPTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKKAPKELAPT-T 763
Db 1222 SPAPSPAPSP-APSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1276
QY 764 TKGPTSTTSKAPAPTTPKETAPTTPKEPAPTTTPKAPAPTTPETPTPT-TSEVSTPTTKE 822
Db 1277 SPAPSPAPSPAPSPSP-APSPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1333
QY 823 PTTIHKSPDESTPELSAETPKALE-----NSPKEGV---PTT 858
Db 1334 PLPPSPAPPLVPVPPAGPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 1393
QY 859 KTP-----AATKPE----- 867
Db 1394 PTFPLAPLPDCTLLAQALLSTPDAANSVTVSAGLFTSVAPSTPELLASFCTVCSCQ 1453
QY 868 MTTTA----- 872
Db 1454 LTATLSLVGSSRGNTNGSSGNGYNGGDAAIQRIAGNSTDRGSSSGSSGSSSW 1513
QY 873 -----KDKTERDL-----RTTP----- 895
Db 1514 GPGTETAEAWADVQDGTYLQLSIGGVYTRTVVVDTRTPSPVSGNVTLSANRIKQEPS 1573
QY 886 -----ETTAAPKMT----- 895
Db 1574 AVGEASLNALGSKQAMLLTISFSEVPADPDPAASLIIVTGALVAEWAAADKMTFYVLAMT 1633
QY 896 --KETATTTEKTESKIT----- 911
Db 1634 LPAELVATAAGSSSSGTSRSGNGTAAAAAAPPAGTTGRRALQQAAPPPPA 1693
QY 912 ----- 911

```

Wed Oct 13 12:38:27 2004

seq1-g.rup

```
Db 1694 SGSSSLSGAATANQQQRHVHFLLPATVADAARNGRNDLSVLTNDVAVPAVGEA 1753
QY 912 -ATTGQVTSITQDTTFFPKITTKITTKTTLAPKV- 943
Db 1754 LAITARTATYP-----AVATITLVAASSSPAQAIKRAKSLLGSGYHILQMLTMSL 1806
QY 944 -----TTKTKTITTE----- 953
Db 1807 YLASRGVGREYGEVAFKVAVLGKGNLGPFAEAMPTNEKEVTAABQARQVGGLDLPIG 1866
QY 954 ---IMNKPETAKPKDRATSKATTPKPKP- 981
Db 1867 NDLLGSSNTTASGSSGSSSSNSP-PRRPPPPAAGSTGLLESNADAPPPPLAVATPA 1925
QY 982 -----TKAPKKTSTKPKTMRVVRKPTTTPRKMST--MPELN- 1020
Db 1926 LPAPLPSTIAAATAAP-----PRLPSP--PPAVGSGSTGLPRRLHMQML 1970
QY 1021 -----PISRIAEAMLOTTTRPNQTPNSKLVE----- 1046
Db 1971 QPPAAVAAPPPPPASSSALVLPSPPPPPPPSLLIQASATYVSDMQDILLYTLVVAAM 2030
QY 1047 -----VNPKSEDAGABETPHMLLRPHVFMPEVTDMDYLPVRPN 1087
Db 2031 LITAVAGRLIAVLYRLLSPE-----PPLFLA-----FPRLE- 2064
QY 1088 QGIIINPMLSDETNICNKPVDGLTTLRNGTLVAFRGH- 1125
Db 2065 -----TTIAGLLVALTYSCMALGGPAADMHGSRTAAYCV 2100
QY 1126 -----YFWMLSPPSPSPARRITEVWGP- 1160
Db 2101 LTIAVVPYAAFLWLA-----LARAMVPOFTLVEPMTTSSYSPSPSAFERADR 2150
QY 1161 -----CEGKTFKKDSQY--- 1173
Db 2151 TSPPRSTVSIATAADIGTATCGGGGAAGSAGCADVGAMAAATVAAGADVPSDDRYARG 2210
QY 1174 --WR--FTNDIKDAGYKPIFKFGGLTGQIQAALSTAKYXNWPESVYFFKGGSIQQYIY 1230
Db 2211 PHWKQF---DGVLPPTVASGVGGGGVVL- 2238
QY 1231 KOEPVQKCPGRPALNYP-----VYGEMTQVRRREREEA 1264
Db 2239 --PPLVALFGSGRHLPLPLPAAGAATAAGGAATAPGSPRGAAGE- 2284
QY 1265 IGPSOTHTIRIQYSPARLAYQDKGVLHNEVKVSIILWR- 1301
Db 2285 ---DTH---QFGP-----H---WRKFSVPDDGAATAGVGVDAAGGA 2317
QY 1302 ----- 1301
Db 2318 GGGGGRSGSGTGDVRFANGTRTPSDDGAKGRSSHGGGNGNAMS SGAASFGASAGG 2377
QY 1302 -----GLP-----NVVTSALS- 1313
Db 2378 GAAAASRLSGRSGPGDYNAMSGAAMRGAGASRFGPAGSRAGSPPEGAGGYAGGA 2437
QY 1314 -----PNIRKP-- 1319
Db 2438 NAMFAGTAGRAGAISPAHDGSRPFGSGNEGSRQWYSGNNAMTAGAVGVGGVRRPAG 2497
QY 1320 -----DGVDYAFSKDQYNIIDVPS-----RTARAI- 1345
Db 2498 SLRNASDG-----DAPPLPAPGTVPPLFANLLTSGADVRSRRALNAEAVA 2543
QY 1346 -----TTRSG 1350
Db 2544 AAAAGGGGGGTSRSG 2559
```

RESULT 15
Q6SSE8

```
ID Q6SSE8 PRELIMINARY; PRT; 3889 AA.
AC Q6SSE8;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Minus agglutinin.
GN Name=SADI;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RZ SEQUENCE FROM N. A.
RC STRAIN=CC-621;
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.,
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY450929; AAC07042.1; -.
DR InterPro; IPR006315; Autotransporter.
DR InterPro; IPR008985; ConA like lec.gl.
DR InterPro; IPR003882; Pistil extensin.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR TIGRFAMs; TIGR01414; autotrans_barli.1.
DR SEQUENCE 3889 AA; 389223 MW; 4D5B444D5507214A CRC64;
SQ
Query Match 16.2%; Score 1184.8; DB 2; Length 3889;
Best Local Similarity 12.1%; Pred. No. 3e-15;
Matches 438; Conservative 187; Mismatches 541; Indels 2452; Gaps 92;
QY 1 MAWKTLPIYLLLLSVFVIQOVSS-----QQLS----- 28
Db 22 LAFATL-----FPLDLASTQTTGYPMNATEDNEQHKGLLAFILSGDTSFWRSP 71
QY 29 SCAGRCGG-----YSRATCNCDYNCQHYE-----CC----- 57
Db 72 EVATRLGEGTAPWCRCISNCQIQTIFSDQCAPDCESRYCEPGGAALGSENTCCALSLL 131
QY 58 -----PDFKRVCT- 65
Db 132 DOTYASAQPPSSTQPCWCTYPGMGARPPSVCDNFVFAARGTTPAGADDPDLAVSCSS 191
QY 66 ----- 65
Db 192 GTVPTYISGTRYQRNDTVYRIMHNGAVTNPANVTRNQVSIKLRHSAMWHHPSPVNIISP 251
QY 66 -----AELSC-----KGRCF----- 75
Db 252 PFFSLVSELACLPLEIEFIEDVALRADYSVLMAQPNVTLIDNGKVFNLDPSTYPFNVTLG 311
QY 76 -----ESFERG-----REC-----D 85
Db 312 YNNVFSMLIEPSFMLHSSVSVKASSFNPGALRNAALRTCNFAPFTELHAMRWTOQALD 371
QY 86 CDAQ----- 89
Db 372 WSAEYQAQALSYVRDGSQRLPRGWLPPVIEPVVSARLLLPTLRLRTRTRDEQHGAVQT 431
QY 90 -----CK 91
Db 432 SRPLITGPLPGEWALLRNLEYLDLSDMETGALVGPISPTLWMLMHLRVINMTGHEHFGR 491
QY 92 KYDK-----PD-----Y--ESFC 104
Db 492 DWHKIISQIRMYRAATHEPNLNVPHYYPGWGNGNMTYRINISVYDLSGHGWYDEV 551
QY 96 -----CC----- 104
Db 552 TTEAGFEVIAHPHGQCCMKWSQTIKDNNEYILYPGSRFGNNVQDELYGGFYQDEWC 611
QY 105 AEVHNPTS-----PPSSKKAP----- 120
Db 612 ----EFTSPQPPPPPPAPPSPSPPTTBDVPPMPPSPSPAFVMPAPPAPPPIPPASPLT 667
```

QY 121 ----- 120
Db 668 PAAPRRPPLPTWPGKWGAMPFRPPIPRPRPRPPPLPPSPPLPPVTPSPRRPPPK 727
QY 121 ----- 120
Db 728 SPFPKPSPPRSPRRPRLPLPPSPPPPPPLPNPSPAPPPPPPPPSIPPPSPG 787
QY 121 -----PPSGA--SOTIKSTTKRSPKPNKKKKVIES----- 151
Db 788 PPSPPPPSAPPSAAPPSPMPPSPAPPPDPSPKPPSPVPPSPPLPPSPPPSPVPPSP 847
QY 152 -----BEITEVDKNKRNKKTDPKPPVVDEAGSLDNGDFKVTTPDTSTTOHKNVSTSP 207
Db 848 PASPPTSPAPSPPPPPSPPPSPAPP-----SPPPSPPPSPAPSP 891
QY 208 KITTAKPINPRSLPNSDTSKETSITVKNKTTTNTKSTTDGKTKTSKET 267
Db 892 PLPSPPSPAPPLPP----- 907
QY 268 QSIKTSKDLAPTCKVLAKPTPKAETTTKGPALTTTPKEPTTTTPKEPASTTPKEPTTT 327
Db 908 ----PPSPPPSPAPSPPPSPPPSPAPSPPPSPPPSP-APPSPPSPPPSPAP 962
QY 328 IKSAPTTKEPAPTTTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 387
Db 963 PLPPPPSPPPSPAPSPPPSPPPSPAPLPPPPSPPPSPAPSPPPSPPPSPAP 1022
QY 388 TPKEPAPTTKPPAPTTKKEPAP---TTTKEPTTTTKEPAPTTTKEPAPTTTKEPAPT 443
Db 1023 SPFPSP-EPSPAPSPPPSPPPSPPPSPAPSPAPSPPPSPAPSPPPSPAPSPPPSP-A 1080
QY 444 PKAPAPTPK--EPAPTTKKEPAP---TTTKEPSTTPKEPAPTTTKEPAPT 497
Db 1081 PPSAPSPPPSPAPSPPPSPPPSPPPSPPPSPPPSPPPSPPPSPPPSPPPSP 1140
QY 498 TTKSAPTTKEPSTTTTKEPAPTTTKEPAPTTTKEPAPT-----TPKEPAPTTTKEPAP 551
Db 1141 PPSPPSPPPSPAPSP--PPSPSPPPSPPPSPPPSPPPSPPPSPAPSPAPLPLPPSP 1198
QY 552 TTTKAPAPKAPAPTTKAPTTKAPTTKAPTTTKEPAPTTTKEPAPTTTKEPAPTPELAP 611
Db 1199 HTQSPSP-VPPSPAPSPPPSPPPSPPPSPPPSP-APQAPSSPP-PPQPTAPTAP 1255
QY 612 EPTTTPKEPAPTPK-----AAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTP 727
Db 1256 PPSPPSPAPSPPPSPPPSPAPQPPSPPTTKEPAPTTTKEPAPTTTKEPAPTTPK 667
QY 668 GTAPTTKEPAPTTKPPAPKELAPTTTKEPSTTSDKAPTTTKEPAPTTP 727
Db 1314 -----SPAPSVFPPSP--APSPMPPS-----PAPLAPQPPSP-TTPSPAPVP 1355
QY 728 KEAPTTTGTAPTTTKEPAPTTTKEPAPKELAPTTTKEPSTTSDKAPTTTKEPAPTTP 787
Db 1356 PPSPP 787
QY 788 PKEPAPTPK----KPAPTTTTPPTTSEVSTTTTKEPPTTTHKSPDESSTPELSAEP 843
Db 1407 PLQSPFPPPSPPPSPAPPPSP-----SPSTTSPSPAPLAPAPVPVPMAPQPPS 1460
QY 844 KALENSKPEGVPTTKTPAATPE----- 867
Db 1461 PPLPSPPPPPPPSPPTTASPOFAPAAVLDGSAATRTSFVASSRSGAFYIAVAVAS 1520
QY 868 ----- 867
Db 1521 SPSYQVCCCELSYAVLDGASQQVIVIPSSGSSSTAGSPVAVTSSVSTPAGAGLNG 1580
QY 868 -----MTTAKDKTTERDLR----- 882
Db 1581 THGSTARRALWVEATASSGPAAGVGARHILLATANTSTLEGLLATGRSAAAGMG 1640
QY 883 -----TTPETT----- 888

Db 1641 MSRLQVVDVQTGCLDPTVTAAPPTGTSPNTTSGAGEAGSGTTRYSSMGAGSGGLDAWR 1700
QY 889 ----- 888
Db 1701 LTPGATGDYLLRLKVADQEWVWVSVDIDPPRAAQQLLIARRTTGGSSSSNSTSGSALA 1760
QY 889 ----- 888
Db 1761 AAEDEVEVNAHAAGAATAVSAQAUVRLMMVIAVSEVPQPFSLTSALSLGARLL 1820
QY 889 -----TAAPKM-----TKETATTEKTTESKITATTTQVITTT-- 922
Db 1821 STQCFASATAAAEAAAAGTVDASTPPGSDASSAATVAPAAIAFVSGSTSGTSSNSG 1880
QY 923 -----QDTTFFKITTLKTTT-----LAPKVTTTKTTTTTEIMN 956
Db 1881 SAYQSCVAVLFAEQDATPELLPPGLTDMHGNINAEPLILSVNLTASADSLSTVERAG 1940
QY 957 KPEETA-----KPKOR 967
Db 1941 APVAAAAGGVFAAFTSASASFLSAFSSRSLLOSQYHIQMLAMSSSLASPGISPAF 2000
QY 968 -----ATNSKA----- 973
Db 2001 RISRYLWSLIGIOGNIPLLDGAFSGSAAAGSGSSSSSSGGLGDVDVAAVALDRL 2060
QY 974 --TTPKP-----QKP----- 981
Db 2061 QLSVPPPLPAAGDAASQAQPPANLPPSPSASQLVADGSTALAGRRSRSLVQAAAPVADSP 2120
QY 982 --TKAPKKP---TSTKKPKTMPR-VRKPKTTP-----INP----- 1007
Db 2121 PFTQAPPTAPFGTGTGAAPAPPPAPPPQSPSTPPPPPPPPALSUGDRDVLVWLQVGA 2180
QY 1008 -----TPRMTSTMP----- 1021
Db 2181 IGSNSSGGSATSASVAGYSLGGAASPRGVVVLGQLGQLGPGGSAPPMPQG 2240
QY 1022 -----TSRIAEAMLOTTTTPN----- 1037
Db 2241 SGSGGQSSGASVATDTTTHNVQDLYLTIAIALLMVALVAHLLVIGLRLAVMYDVC 2300
QY 1038 -----QTPNSKL----- 1045
Db 2301 GAAESGVEGLHPLVRFPRAEMVLGGLLLVALTFYSALTLSGAASPRWGDNTAAGRLI 2360
QY 1046 ----- 1045
Db 2361 VLAVLVVYGLLLMMLTVCRWYLQEEVDHYMLGPHWQAFDGVIPGGAGAGDGGHGA 2420
QY 1046 -----EYNPKSEDAGGAGET-----PHML----- 1066
Db 2421 SALPAGVFGTGGGGVAFACQAEPPSGDGGGEGDGYGLGPHWALAPAGAKTLDYE 2480
QY 1067 ----- 1066
Db 2481 SIASPAAGAGSAGATRTDAAGSTVGGAGSKGRPRTLRSTDTVTAGGAAAGTAAATAA 2540
QY 1067 ----- 1066
Db 2541 AAAAGAPAGAGDVGAAFORRYGEADNHDGSDYDGEDGDLDMAPRRRPPADGVDA 2600
QY 1067 -----RPHV----- 1070
Db 2601 AAPATAARKMAGSDDGAAGERRRANPHARKSADGAGDAGDPQAGRPLLRRGVSYGDN 2660
QY 1071 -----FMPEVT----- 1076
Db 2661 LLTAGANTRGSAIRSTPCPVADDDARSFRFRNTVYAGDNALTSGAASAGAAPSPTRTQ 2720
QY 1077 PDMYDLPRVFNQ----- 1089
||| | | | | |

Wed Oct 13 12:38:27 2004

seq1-g.rup

```

Db      2721 PDSPQAPLSPRRGAAMGNAYTAGASRFGGPRTRADSPRAASPPPEPASPLSGDAGALS 2780
Qy      1090 -----III 1092
Db      2781 EVGFNANWGGSRPRSDPTWEDEAAAIQODAADPAVEANRERKQAGSVEEQQVAI 2840
Qy      1093 NPMLSDEINI-----CNGKPDVGLTTLRNGTLVAFRGHYFWMLSPSPSPS 1137
Db      2841 NPMVPRASVLERPDVAASLAASGAPPQGTVEL-----PP- 2879
Qy      1138 PARRITEVWGIP-----SPIDVTFRNCCEGKTFKFDQYMRFTNDI----- 1180
Db      2880 -----ITEATG-ENGLGSSPRD-----GRYMTTASVSHGRDTSRLHLHPMARPSLGA 2925
Qy      1181 -----KDAGYKPIFKG----- 1192
Db      2926 SLVPASATNGSGNSGNSGREGSGSSTSGGAPAAAMPFRVGGAPSNASAAAAA 2985
Qy      1193 -----FGHITGOIVAALSTAKY-----KNWPE-----SVYFFKRGSIQ----- 1226
Db      2986 AQPAAHQDLFGELDTTTLTKMTADAGAGAPPENLPSMMRLGSLGLVVRGSGNSGAAQV 3045
Qy      1227 -----QYIY----- 1230
Db      3046 APASPLPAKAPSPPAARPPGAVTTNAAVPMPPQKVERSPSEPANARAQGAAPPAA 3105
Qy      1231 -----KQEPVQK-----PQRRPAL----- 1245
Db      3106 AAVANGGRELKQQPAVQAVSSSTSSIRSNSSGSGNPLARALRYMSVPKGQAPSTASNA 3165
Qy      1246 ----- 1245
Db      3166 APAIAEAEPAPPSRHTAWGGSLQPPAAAAAGAPPPIAPQSPMRLPMPPPGAGGAGG 3225
Qy      1246 -----NYPVYGMTQVRRRFFERAI-----GP----- 1267
Db      3226 AGAKAGGFTSPQASIVMESSPAAGPVMPARRRRNRRAIMDEDDLVAEAGGAGPVAE 3285
Qy      1268 -----SQT-----H 1271
Db      3286 AEAPDSAGHYGGWEDEPRLLLEAPRSPVMPPPPLRAGGAARFSGSSTGMSAAAVSVH 3345
Qy      1272 TIRIQYSPARLAYQDKGVHNEVKVSIWLRG-----LPNVVT----- 1308
Db      3346 S-----YSPSSA-----RGGGTAAGHYMTYNPLSLPSAVSAAAEAAA 3384
Qy      1309 ---SAISLP---NIRKPDGYDYA-----PSKDQYNNIDVPSRTARAI- 1345
Db      3385 VTPSALASPLSLSLPRLPRGRSQQAAGSPRSVLLILQSRSTFA-----PNPTAMAIL 3435
Qy      1346 -----TTRSGQT 1352
Db      3436 GTRGGGGVGGCSTRSLQS 3453

```

Search completed: October 13, 2004, 12:20:00
Job time : 199.642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 27.7222 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-G
Perfect score: 7223
Sequence: 1 MAWKLPYILLLLSVFVIQ.....ARAITTRSGQTLSKVWVNC P 1361

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 79: *
2: PIR: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|-----------------------------|
| 1 | 1395.9 | 19.1 | 3020 | 2 A439332 | ----- mucin 2 precursor, |
| 2 | 1179 | 16.1 | 1664 | 2 T18262 | S-layer protein - |
| 3 | 1061.2 | 14.5 | 7962 | 2 I38346 | elastic titin - hu |
| 4 | 1050.9 | 14.4 | 2187 | 2 T30826 | nascent polypeptid |
| 5 | 1048.4 | 14.3 | 3570 | 2 T45025 | mucin MUC5B, trach |
| 6 | 1027.3 | 14.0 | 1489 | 2 T31108 | cyst germination s |
| 7 | 1018.4 | 13.9 | 1274 | 2 T16251 | hypothetical prote |
| 8 | 983.3 | 13.4 | 3507 | 2 T34513 | hypothetical prote |
| 9 | 971.4 | 13.3 | 6642 | 2 T29757 | protein UNC-89 - C |
| 10 | 971.3 | 13.3 | 1367 | 1 S48478 | glucan 1,4-alpha-g |
| 11 | 931.3 | 12.7 | 4135 | 1 S42629 | tenascin-X - bovin |
| 12 | 917.8 | 12.5 | 1188 | 2 S49915 | extensin-like prot |
| 13 | 916.8 | 12.5 | 5762 | 2 A1819 | proline-rich pepti |
| 14 | 882.5 | 12.1 | 1229 | 2 T25697 | hypothetical prote |
| 15 | 881.4 | 12.0 | 2897 | 2 B48666 | cell proliferation |
| 16 | 875.7 | 12.0 | 1344 | 1 A35175 | mucin 1 precursor, |
| 17 | 869.3 | 11.9 | 3256 | 2 A48666 | cell proliferation |
| 18 | 862.6 | 11.8 | 1151 | 2 T18353 | high molecular mas |
| 19 | 860 | 11.7 | 5262 | 2 T03454 | ALR protein - huma |
| 20 | 850.6 | 11.6 | 3942 | 2 T42730 | Bassoon protein - |
| 21 | 836.1 | 11.4 | 2232 | 2 T34434 | hypothetical prote |
| 22 | 829.6 | 11.3 | 3938 | 2 T42761 | Bassoon protein - |
| 23 | 828.4 | 11.3 | 2142 | 2 B35098 | MHC class III hist |
| 24 | 823.6 | 11.2 | 4006 | 2 T09070 | probable tenascin |
| 25 | 822.4 | 11.2 | 4957 | 2 T03455 | ALR protein - huma |
| 26 | 820.9 | 11.2 | 1832 | 2 T31113 | mucin-like glycopr |
| 27 | 817.3 | 11.2 | 4548 | 1 S00657 | apoptoteina(a) (BC |
| 28 | 809.9 | 11.1 | 1872 | 2 S36132 | MHC class III hist |
| 29 | 802.9 | 11.0 | 3381 | 2 T42389 | versican precursor |

ALIGNMENTS

RESULT 1

A439332
mucin 2 precursor, intestinal - human (fragments)
N: Alternate names: mucin SMUC-41
C: Species: Homo sapiens (man)
C: Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004
C: Accession: A49963; A45106; A43932; B33532; B61257; P00328; P00329
R: Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A: Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A: Reference number: A49963; MUID:94132002; PMID:8300571
A: Accession: A49963
A: Molecule type: mRNA
A: Residues: 1-639 <GU1>
A: Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998
R: Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A: Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A: Reference number: A45106; MUID:93016075; PMID:1400449
A: Accession: A45106
A: Status: not compared with conceptual translation
A: Molecule type: mRNA
A: Residues: 626-1895 <GU2>
A: Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396
A: Note: sequence extracted from NCBI backbone (NCBIP:116706)
A: Accession: B45106
A: Status: not compared with conceptual translation
A: Molecule type: mRNA
A: Residues: 2037-3020 <GU3>
A: Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398
A: Experimental source: colon
R: Toribara, N.W.; Gum Jr., J.R.; Culhane, P.O.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1008-1013, 1991
A: Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A: Reference number: A43932; MUID:91358717; PMID:1885763
A: Accession: A43932
A: Molecule type: DNA
A: Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A: Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864
A: Note: sequence inconsistent with the nucleotide translation
A: Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R: Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A: Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A: Reference number: A33532; MUID:89197956; PMID:2703501
A: Accession: B33532
A: Molecule type: mRNA
A: Residues: 1916-2193 <GU4>
A: Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874
A: Experimental source: intestine
R: Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

30 799.9 10.9 2225 2 T26063 hypothetical prote
31 799.3 10.9 1870 2 S37671 MHC class III hist
32 792.6 10.8 2774 2 A43359 microtubule-associ
33 782 10.7 1611 2 T38236 hypothetical prote
34 777.8 10.6 5105 2 T32650 hypothetical prote
35 777.3 10.6 990 2 I51618 nucleolar phosphop
36 774.4 10.6 761 2 C84672 hypothetical prote
37 773.4 10.6 1630 2 A53577 ascites sialoglyco
38 769 10.5 3566 1 A40701 tenascin-X precurs
39 766.4 10.5 4667 2 T20774 hypothetical prote
40 765.8 10.5 4549 2 T20771 hypothetical prote
41 760.7 10.4 5170 2 T15348 hypothetical prote
42 757.2 10.3 924 2 S27923 gene IF3 protein -
43 756.2 10.3 3968 2 A44265 trithorax homolog
44 755.2 10.3 1777 2 T34369 hypothetical prote
45 754.5 10.3 13055 2 T16580 hypothetical prote

| | | | | |
|--|----|------|---|------|
| J. Clin. Invest. 87, 77-82, 1991 | Db | 1076 | VHDSCTDGGDCFCFSAVASAQAQCTKEGACVFWRTPDLCPICFYVNPPECEWHYE | 1135 |
| A:Title: Human bronchus, and intestine express the same mucin gene. | | | | |
| A:Reference number: A61257; MUID:91086481; PMID:1985113 | QY | 208 | -----KIITAKPIN----- | 216 |
| A:Accession: A61257 | Db | 1136 | PCGNRSFETCRTINGIHNSISVSYLEGCVPRCPKDRPIYEEDLKCCVTADKCGVEDTH | 1195 |
| A:Molecule type: mRNA | QY | 217 | --PRPSLP----- | 230 |
| A:Residues: 'T',1925-1948,'TTS',1952-1954 <JAN> | Db | 1196 | YPPGASVTEETCKSCVCTNSQVVCVRPEEGKILNQDGAFCYWEICGPNGTVEKHFN | 1255 |
| A:Experimental source: bronchus | QY | 231 | TSLTVNKET-----TVETKETTNNKOTSTDGKEKITSAK----- | 265 |
| R.Xu, G. Huan, L. Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992 | Db | 1256 | CSITTRPSLTITFTTPTTPTSTTTTTTTTTTSSVLSTPKLCLLWSWDINEDHPS | 1315 |
| A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-b | QY | 266 | -----ETQSIEKTSKAD----- | 277 |
| A:Reference number: PQ0328; MUID:92198477; PMID:1550588 | Db | 1316 | SGSDGDRPFQVCGAPEDIICRSVKDQPHLSLEHQHGKQVQCDVSFGICKNEDQFGNGP | 1375 |
| A:Accession: PQ0328 | QY | 278 | -----LAPTSKVLAKPTPKAETTTKGPAITTPKBPPTTTPKBPAS | 317 |
| A:Molecule type: mRNA | Db | 1376 | FGLCYDYKIRVNCWPMDCIITPSPPTTTPSPPTTTLPTTTPSPPTTTLPTTTPPT | 1435 |
| A:Residues: 2328-2468 <XUG> | QY | 318 | TTPKETPTTIKSAAP--TTPKEPAPTTTKSAPTTPKEPAPTTTKKEPAPTTTKE | 376 |
| A:Cross-references: GB:M8523 | Db | 1436 | TTTSPPTITTTPLPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPT | 1493 |
| A:Experimental source: small intestine | QY | 377 | PAPTTTKSAAP--TTP-----KEPAPTTPKAPTTTPKEPAPTTTKEPAPTTTKEPAPTT | 427 |
| A:Accession: PQ0329 | Db | 1494 | PPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPST | 1553 |
| A:Molecule type: protein | QY | 428 | TKBPAPTTKEPAPTTAKKPAPTTTPKEPAPTTTPKEPAPTTTKBPSPPTTKBPAPTTT | 487 |
| A:Residues: 2328-2342,'K',2344-2354 <XUG1> | Db | 1554 | TTLPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTP | 1613 |
| C:Genetics: | QY | 488 | PTTTPKEPAPTTTKSAAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT | 546 |
| A:Gene: GDB:MUC2 | Db | 1614 | PTTTPSP--PTT--PTT-----PTSTTTLPTTTPSPPTTTPPTTTPSPPTTTPPTTTP | 1665 |
| A:Cross-references: GDB:120203; OMIM:158370 | QY | 547 | KEPAPTTTKBPAPTTAPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT | 606 |
| A:Map position: lip15.5-lip15.5 | Db | 1666 | SPPTTTPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPSS | 1717 |
| C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von | QY | 607 | PTTPEEPTPTTPEEAPTTPKAAAPNTPKEPA-----PTTPKEPAPTTTPKEPAPTT | 658 |
| C:Keywords: glycoprotein; intestine; tandem repeat | Db | 1718 | TTTP--SPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP | 1769 |
| F;2766-2834/Domain: von Willebrand factor type C repeat homology <WVC> | QY | 659 | -----KETAPTP-----KG----- | 668 |
| Query Match 19.1%; Score 1395.9; DB 2; Length 3020; | Db | 1770 | FSPESTTTPPTTTPCPLCNWTGWLDSGKPNFHKGGDTLIGDVCVPGWAANISCRATMYP | 1829 |
| Best Local Similarity 18.7%; Pred. No. 3.8e-21; | QY | 669 | ----- | 678 |
| Matches 492; Conservative 124; Mismatches 585; Indels 1431; Gaps 90; | Db | 1830 | DVPIGQLGQTVVCDVSVGLICKNEDQKPGGVIPIMAFCLNVEINVQCCCVQPTTM--T | 1886 |
| QY 6 LPVILLLLSVFVIQVSSQDLSSCAGRCG-----EG-----YSR 40 | QY | 679 | PTTPKKAPKAPLAPTT--TKEPTSTSDKP-----APT--TPKGAPTTTPKEPAPTTPK | 731 |
| Db 502 LQVQLAPVQLFVTLDAQSQ--GQVQLCGNFENGLEDGDFKTASGLVEATGAGFANTWKA 559 | Db | 1887 | TTTTENPTTPTTTTTTTTPTTPTTPTTSTQSPNGLOAPTPTTSTTTTPTTPTTPTTPT | 1946 |
| QY 41 DATCN-----CDV-- 48 | QY | 732 | PTTPKGAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT | 790 |
| Db 560 QSTCHDKLDLWDDPCSLNIBSANYAHCWSLLKKTETPFGRCHSAVDPAEYKCKYDTC 619 | Db | 1947 | TTPTTTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT | 2005 |
| QY 49 NCQHYMEC---CPDFKRVCTA-----ELSK 71 | QY | 791 | PAPTTTPKP-----APTTPETPPTTSEVSTPTTK-----EPTTIHKSPDESTEP | 838 |
| Db 620 NCQNNEDCLCAALSSYARACTAKGVLWGRHVNCNDKVGSCPNQVFLYNLTTCQCTCR 679 | Db | 2006 | PTTPTQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT | 2065 |
| QY 72 ---CRCPESF-----ERGR-----ECDC-----DAQCKYDK 95 | QY | 839 | ABPTKALENSKEPGVPTTKTAAKEMTTAKDKTTERDLRTPPTTTTAAKPMKTKET | 898 |
| Db 680 SLSEADSHCLEGFAPVDGCGCPDHTFLDEKGRCVPLAKCSYHRGLYLEAGDVVVOEER 739 | Db | 2066 | RSTSSPLTESLTLLSTLPPAIEMTSTAPPSTTAPTSTSGGHTLSPPTTTPSPPTTPT | 2125 |
| QY 96 C----- 96 | QY | 899 | ATTTEKTTESKLTATTTQVSTTQDTTPFKITT--LKTTTLAPKVTTTKKTTITTEIM | 955 |
| Db 740 CVCRDGRLHCFQIRLIGQSCCTAPKIHMDCSNLTALATSKPRALSCQTLAAGYYHTECVSG 799 | QY | | | |
| QY 97 ---CPD-----YESFCAEVHNPTSPSSKKAP----- 120 | QY | | | |
| Db 800 CVCPEGLMDGRCGVKEKPCVHNNDLYSGAKIKYDCNTCTCKRGWVCTQAVCHGT 859 | QY | | | |
| QY 121 -----PPSG-----ASQT 128 | QY | | | |
| Db 860 CSIYSGHYITFDGKYDFDGHCSYVAVQDYCGQNSILGSFSIITENVPCTGTGVTCSKA 919 | QY | | | |
| QY 129 IKSTTKRSPKPNKKTKKVIKVESEILTEVKDNKNT----- 165 | QY | | | |
| Db 920 IKIFMR-----TELKDKHVRVIOQDEGHVAVTTREVQYLVV 960 | QY | | | |
| QY 166 -----KKKPTPKPPVVDEAGSL-----DNGDFKV----- 190 | QY | | | |
| Db 961 ESSTGIIVDKRTTVFKLAPSYKGV-----CGLCGNFDRSNNDFTTRDHMVVSSSL 1015 | QY | | | |
| QY 191 -----TTPDSTTQ-----HNKYSTP----- 207 | QY | | | |
| Db 1016 DFGNSWKEAPTCTPDVSTNPEPCSLNPHRRSMAEKQCSILKSSVFSICHSKVDPKPFYEAC 1075 | QY | | | |
| QY 208 ----- 207 | QY | | | |

Db 2126 GTTT--GSSAAPTPTVQTTTSAWTPPTPLSTFSLIIRTKLGRYPSSVLICCVLNDTY 2183
QY 956 NKPEE-----TAKPKDRATNSKATTPKQKTKAPK 986
Db 2184 YAPGEEVNGTYGDCYFVNCSLCTSLFYWNWSCFTSPPTPSK-STPTSKFSSTPS 2242
QY 987 KPTSTKTKTMRVRKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPNCOTPSKLVE 1046
Db 2243 KPTGCTKPECFDFDPPR-----QENETWMLCOCFM-ATCKYNNTVIIVKVE 2288
QY 1047 VNP-----KSEDAGGA-----EGTPHML 1065
Db 2289 CEPMPMTCNGLQPVREDDDGCGWHWCDCYCTGWDGPHVYVTPDGLYYSQGNCTVVL 2348
QY 1066 LRPHVMEVTPDMYDLPRVFNQGIINPMLSDENICNGKPVQGLTLRN----- 1116
Db 2349 VE-----EISPSVD-----NFGYVINDYHCD-----PNDKVSCPTRLIVRHETOE 2388
QY 1117 -----GTLVAFRG----- 1124
Db 2389 VLIKTVHMPMQVQVQVNRQAVLPYKYGLEVYQSGINVVVDIPELGLVLSYNGLSFSV 2448
QY 1125 ---HYF-----WMLGFFSPP----- 1136
Db 2449 RLPYHRFGNNTKGCGTCTNTSDCILPSGEIVSNCEAAADQWLNDPSKPHCHSST 2508
QY 1137 -----SPARRITEVWGIPSPI-----DTVFTRCN-----CEGKTF 1166
Db 2509 TKRPAVTVPGGKTPHKDCT-----PSPQLQIKDSLFAQCHALVPPQHYDCAVDFSC 2563
QY 1167 FFKDSQV-----WR----- 1175
Db 2564 FMPGSSLECASLOAYALCAQONICLDNRNHTGACLVECPSHREYQACGPAEPTCKSS 2623
QY 1176 -----FTNDIKDAGY--PKPIFKGFGGLTQIIVAAALSTAK 1208
Db 2624 SSQNNNTLVGECFCPEGTMVYAPGDVVCVTKCGCVGPDNVPRBF----- 2669
QY 1209 YKNWPSVYFFK-----RGGS--IQQYTYKQEPVQKC----- 1238
Db 2670 -----EHFEDCKNCVLEGGSLICQPKCSQKPVTHCVEDGTYLATEVNPADTCC 2721
QY 1239 -----PGR-----RPALNYP 1248
Db 2722 NITVCKNTSLCKBPSVCPGLFEVSKMVPGRCCPFYWCESKGVGVHGAEBQV--GSP 2779
QY 1249 VYGEWTQ----- 1257
Db 2780 VYSSKQDCVCTDKVDNNTLLNVIACHTVPCNTSCSPGFELMEAPGECCKCEQTHCIK 2839
QY 1258 RRRFERAIGPSOHTIRIQYSPARLAYQDKG-----VLEHNVKVSILWGLPNVNTS 1309
Db 2840 R-----PDNQHV-----LKPGDKSDPKNNCTFPSCVKIHNQLISSV-----S 2878
QY 1310 ATSLN-----IRKPGDYDYAFSKQOYINIDVPSRTARAITR----- 1348
Db 2879 NITCFNFASICPSITF-----MPNGCKTCTPRNETRVPCTSVPTVTEVS 2926
QY 1349 -----SGQTLISKVWY-----NCP 1361
Db 2927 YAGCTKTVMNHCSGCTFVWYSAKAQALDHSCCKEKTQSREVLVLSCP 2978

RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J. P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulosem
e.

A:Reference number: Z18847; MUID:932099311; PMID:8458832
A:Accession: T18262
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUJ>
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA4784

Query Match 16.1%; Score 1179; DB 2; Length 1664;
Best Local Similarity 20.9%; Pred No. 4.4e-17;
Matches 412; Conservative 138; Mismatches 460; Indels 960; Gaps 85;
QY 4 KTLPIYLLILLISVFIQOVSSQDLSSCAGRCGEGYSDATCNCDYNCHVMECCP----- 58
Db 6 KVLISILLTLL-----IISTTSWNMS-----FAEATPSIEWV 37
QY 59 -----DFKRVCTAELSCKGRCFESFERGREGDCDCAQCKYDKKCCPD 99
Db 38 LDKTEVHVGVITATIKVNNIRKLAGYQLNK----- 69
QY 100 YESFCAEVHNPSTP-----PSSKAPPSPGASQTIKS----- 131
Db 70 ---FDEVLQPVDPATGEEFTDKSMVNVRLVLTNSKYGPTPV-AGNDIKSGIINFATGYN 125
QY 132 --TKRSP-----KPPNKKTK----- 146
Db 126 NLTAAYSSGIDEHTGIIGEIFGVKKKNTSIRFDTLSMPGAISGTSLPDWDAAETIGY 185
QY 147 -----KVIESEI-----TEVK-----DNKN-----RTKKKPT- 170
Db 186 EVIQPLIIVVEAPLKDASVALLDRTKVKVGDIITATIKIENMKNFAGYQLNKYDPTM 245
QY 171 -----PKPPVVDEAGSGLN----- 185
Db 246 LEAIELETSAIAKRTWPVTGVLQSDNVYKKTAVANDVGAGIINFAEAYSNTKYRET 305
QY 186 -----GDFKVTTPTD-----STIQNKV 203
Db 306 GVAEBTGIIKIGFRVLKAGSTAIRFEDTTAMPGAIEGYMDWYGENIKYGVVQGEI 365
QY 204 -----STSPKLTAKPINRPSLP-----PNS-----DTSK----- 229
Db 366 VAEGEPEGEETPEPVTPTPTVPTVTEEPVPSLPSYVIMELDKTKVKVGDIITAT 425
QY 230 -----ETSLTVNKETTVEKTTTNNKQISTDCKEKT 261
Db 426 IKIENMKNFAGYQLNKYDPTMLEALELETGSAIAKRTWPVTGGV-----LQSDNYGKT 480
QY 262 TS-----AKETQSIKTSKDLAPTSKVLAKPTPKAET 294
Db 481 TAVANDVGAGIINFAEAYSNTKYRETGVAEETGIIGKIFRVLKAGSTAI-----RFD 535
QY 295 TT-----KGPALTTP-----KEPTPT-TPKEPASTTPKEP 323
Db 536 TTAMPGAIEGYMDWYGENIKYGVVQGEIIVAEGBEETPEPVTPTVPTVTEEP 595
QY 324 TPT----- 326
Db 596 VPSELPSYVIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIGIKYDPKVLAEFNIETGD 655
QY 327 ----- 326
Db 656 PIDGTPWAVGGTILKNRDYLTGTVAINNVSKGILNFAAYVYVFDYRECKSEDGTGIIG 715
QY 327 -----TIKSAPTPK-----EPATTTKGAAPTPPK 351
Db 716 NIGFRVLKAEETTRFEELSEMPGSDGTVMLDWYLNRIISGVYVQIPAKAAS-----D 770
QY 352 EPATTTKKEPAPTPKPEPA-----PTTTKEPAPTTTKGAPTTKPEPAPTTKPKPAPTPKPE 407
Db 771 EPIPTDTPDEPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSDPTSD 830
QY 408 PAPT-----TPK-EPPTPT-TPKEPAPT-TPKEPAPTAPKKA--PTTPK 453

Db 831 PTPSDEPTSPDEPTSPDEPTPEEPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPD 890
Qy 454 EPAPT-TPKEPAPTTTKPSPPTPKPAPTTTKSAPTPTTKPAPTTTKSAPTPTKEPSP 512
Db 891 EPTSETPEEPIPTDTPSDEPTSPDEPTSPD-----DEPTSPDEPTSPDEPTSPDEPT 947
Qy 513 TKKEPAPTTPKPAPTTPKKA-----PTPKPAPT-TPKEPAPTTTKP 557
Db 948 DTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 1007
Qy 558 APTAKPAPTTPKTAPTTPKKLPTTPEKAPTTPEKAPT-----TPELAPTTPEEPT 614
Db 1008 EPTSPDEPTSPD-----EPTSPDEPTSPDEPTSPD-----EPTSPDEPT 1046
Qy 615 PT-TPKEPAPTTPKAAPNTPKPAPTTPKEPA-PTTPKEPAPT-TPKETAPTTPKPA 670
Db 1047 PSETPEEPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 1106
Qy 671 PTTLKEPAPTTPKKAPKELAPTTTKEPT-STTSKAPAPTTPKGAPTTPKAPTTPK 729
Db 1107 PTPSDEPTSPD-DEPTSPD-EPTSPDEPTSPDEPTSPDEPTSPDEPTSPD-DE 1161
Qy 730 PAPTTPKGTAPTTLKEPAT-TPKAPKELAPTTTKGPTSTTSKAPAPT---TPKETAP 785
Db 1162 PTPS-----DEPTSPDEPTSPDEPTSPD-----IPTDTPSDEPTSPDEPTSPD-EP 1211
Qy 786 TTPKEPAPT-TPKAPAPTTPEPTTSEVSTTKEPTTIHKSPDEPTSPDEPTSPD 844
Db 1212 TSPDEPTSPDEPTSPDEPTSPD-----EPTSPDEPTSPD-----PSDEPTP-SD 1262
Qy 845 ALENSKPEGVPPTKPAATKPEMTTAKKTTERDLRTTPTTTTAAPKMTKATITTEK 904
Db 1263 --SETPPEP-IPTDTPSDEPTSPDEPTSPD-----EPTSPDEPTSPD-----PS 1319
Qy 905 TTESKITATTTQVTSITQTTTPKLTTLTKTLAPKVTTKITTTTEIMNKPEETAPX 964
Db 1320 TPEEPIPTDTPSDEPTSPDEPTSPD-----PSDEPTP 1354
Qy 965 KDRATNSKATTPKQP-KPTKAPKPTSTKPKTMPRVKPPTTTPPKMTS----- 1014
Db 1355 SDEPTSPDEPTSPDEPTSPDEPT-----PTTTPPTSPITPTSGSGSG 1401
Qy 1015 -----TMDELNPTSRIAEAMQLQTTTRPNQTPNSKLVEVNPKSEDAGGAE 1060
Db 1402 GSGGGGGGGGTVPSTPTPTS-----KPTSTPATPTBIE-EPTSDVPGAIGG 1449
Qy 1061 TPMLLRPHVPMPEVTPMDVLPVPNGIILNPMLSDETNICGKPDVGLTLRNGTLV 1120
Db 1450 EHRAYL-----GYPDG----- 1461
Qy 1121 AFRGHYFWMLSPPSPSPAPRITVWGIPIPIDVFTRCNCEGKTFPPK-----DSQY 1173
Db 1462 SFR-----PERNIT-----RAEAAVIFAKLLGADSESYGAQ 1491
Qy 1174 -----WRTNDIKDA-----GYPKPIFKGGGLT----- 1197
Db 1492 SASPYSLDADTHWAWA-----IKFATSQGLFKGYPDGTGKPDQNTIRAEFATVULFLT 1547
Qy 1198 --GO-IVAALSTAKYN-----WPESVVFRRKGGSI-----QQYIYKQ 1232
Db 1548 VKQEIIMKLAITIDISNPKFDDCVCHWAQE--FIEKLSLGVISYCPDGTFFKPNYIKRS 1605
Qy 1233 EPVQKCPGRRPALNYPVNGEMTQVRRRRFERAIGSQTHITRIQYSPARLAYQDKVLEN 1292
Db 1606 ESV-----ALINRALER--GP----- 1619
Qy 1293 EVKVSILWGLPNVVTSAISLNIKPKDGYDYAF-----SKDQYINID 1336
Db 1620 -----LINGAPKL-----FDDVNB-----SYWAFGDIMDGLDHSYIIE 1652

elastic titin - human (fragment)
C.Species: Homo sapiens (man)
C.Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C.Accession: I38346
R.Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A.Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A.Reference number: A57430; MUID:96026330; PMID:7569978
A.Accession: I38346
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-7962 <RES>
A.Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017426
C.Genetics:
A.Gene: GDB:TTN
A.Cross-references: GDB:127867; OMIM:188840
A.Map position: 2q31-2q31

Query Match 14.5%; Score 1061.2; DB 2; Length 7962;
Best Local Similarity 15.1%; Pred. No. 4.2e-13;
Matches 439; Conservative 129; Mismatches 510; Indels 1828; Gaps 107;
Qy 4 KTLPIYLLLLSVFVIQQVSSQDLSSCAGCGGYSDATCNDY----- 48
Db 5317 QTEPVT-----IKDIENQTVLK-----DNDVFEIDIKINYPEIKLSWKYGT 5359
Qy 49 -----NCO-----HYMECC-----PDFKR----- 62
Db 5360 EKLEPSDKFEISIDGRHILRVKNCQLKQGNVRLVCGPHIASAKLTVIEPAWHLQDV 5419
Qy 63 -----VCT----- 68
Db 5420 TLKGGQTCIMTVQFSVNVKSEWFRNGRILKPOGRHKTEVEHKVHLTIADVAEADGGQY 5479
Qy 69 SCGRCFESPERGECDDAQCKYDK-----CCPDYESFCAE----- 106
Db 5480 TCK---YEDLETSALRIEAEPIQTKRIONIVVSEHQSATFCEVSFDDAIVTWYKGT 5336
Qy 107 -----VHN----- 109
Db 5537 ELTESQYNFRNDGRCHYMIHNVTPDDEGVSVIARLEPRGEARSTAEALYLTTKIEKLE 5596
Qy 110 -----PTSP-----PSSKKAPPPSGASOTIKSTT 133
Db 5597 LKPPDIPDSRVPIPTMBIRAVPPEIIPVAVPVPVLLPTPEEK-PPPKRIEVTKKAV- 5654
Qy 134 KRSPKPNKTKKTVIES-----EETEVKONKNRTKKKTP-----KPPVYDE 178
Db 5655 -----KKDAKKVAKKEMTPREIV-----KKPPPTTLIPAKAPEIIDV 5695
Qy 179 AGSLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINRPSLPNSDTSKTSLSLVNKE 238
Db 5696 -----SSKAEVIMT-----ITRKK-----EVOKE-----KE 5718
Qy 239 TTVTKETTTNKT----- 253
Db 5719 AVYEKKQAVHKEKRVFIESFEPEYDELEVEPYTFPEQYVEEDDEYEEIKVEAKVEH 5778
Qy 254 -----STDGKEKTSKAKETOSI-----EKTSAD 277
Db 5779 EWEEDFEGQYVEREYDEGEBEWEENAYQEREVIQVQVEYVESHERKVPKVPKPK 5838
Qy 278 LAPTSKVLAKPT-PAETTT----- 299
Db 5839 APPPPKVIKPKVIEKIEKTSRRMEBEKVQVTKVPEVSKKIVPKPSRTFPVQEEVTEVKVP 5898
Qy 300 ALTTPK----- 316
Db 5899 AVHTKKMVISBEKMPFASHTESESVTVPEVQKEIVTEEKIHVAISKRVPPPKVPPELPE 5958
Qy 317 STTPKEPTTTI-----KSAPTTPKPAPTTTKSAPTTPKPA-----PTTTKEBPAP 363
Db 5959 KPAPBEVAVPIPKKVPBPAPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6018

Wed Oct 13 12:38:27 2004

seq1-g.rpr

RESULT 4
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB187
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2050/3; 2099/3; 2142/3; 2183/3
A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 14.4%; Score 1050.9; DB 2; Length 2187;
Best Local Similarity 19.24; Pred. No. 4.3e-14; Indels 1151; Gaps 96;
Matches 425; Conservative 161; Mismatches 481;

QY 21 QVSSQDLSSCAGRCGEGYSRDATCNCQHYMECCDFKRVCTABL----- 68
DB 183 QVFSQGLNLKGT-----PCPPDVVRAPPFPHLENPLASVQPLGM 222
QY 69 SC-----KGRCFESFERGRCDDCAQCKYDKCCPDY 100
DB 223 SCPTLSNTSPVKGVPISALTSRLSLNLKG----- 254
QY 101 ESFCAEVHNTSPSSKAPP----- 121
DB 255 -----PVSPPARNTAAPSILAPSTSLGCHLPLLHHSSVDSPIQPCQSLAVSNP 305
QY 122 -----PSGASQTIKS 131
DB 306 TSVCHSGTAASCPERCVPVLPALPSRLLAVDSGAAPSDDKGSSAVTNELCSPGSSNV--A 363
QY 132 TTKESSPPNKKTKVIESIEITEVKNKNTKKP----- 169
DB 364 GTSLSPKASLVPKGSNNALQPLVTVQVPASQTKGLKEIPVSCIGATHALDNPSAISVAPA 423
QY 170 TPKEPVVDEAGSLDNGDFKVTTPDTS-----TOHNKVSTSP-KITTAKPINRPS 220
DB 424 THVPP-----PTSSGL-----VSSKDPASVTSLVVPAAHKQPPAPPASATLGVFVSLPA 474
QY 221 -----LPNSDTS-----KE-----TSLTVNKET 239
DB 475 TEGLNLPISALVNVGAPVSPAQGLPRKDTTLQPLAPIALKESPPSSQSASLSLELSED 534
QY 240 TVETKET-----TTTN-----KQSTDGKE 259
DB 535 TVTKTTGGPAPVVRPAIAGVATTTSLRADSPPAVIRADSCVSPNTVSQPLKRSVTDPM 594
QY 260 KITSKET-----QSIEKT----- 273
DB 595 APRTAKNTAPSTSLVPLASBGCPVASSMALSPQNASVSETALALSPEIKSVFPDPDP 654
QY 274 -----SAKDAPTSKVILAKPT-----PKAETTKGPALT 302
DB 655 LAEISFSNARKVDVAVSHMESSGSRQGHDPDASVTAKGTVVCLADSLDTSVSAKSGSALS 714
QY 303 TKKEP-----TPTTKEPASITPEKPTPTTIKSAPTPK 336
DB 715 GASSPLYPLEVFLPEAGLAVGPKGLNKLSTPTTSPSGAPVPSTGAPSPKGPVPT 774
QY 337 E-----PAPTTTKS-----APTTPKE----- 352

775 ESSISSQVPAIBILPSPOKTPTEVTASRLISAVQSPKVDPIMSDVTPPTSPPKTSATAVPKD 834
QY 353 -----PAPTTTKEP-APTTPKEPAPTTTKEPAPTTTKS--APTTPKE----- 391
DB 835 TSATLSLSKSVPAVTSLSPPKAPVAFSNEA-TIVPTETISLKNALAAATPKETLATSIKP 893
QY 392 ---PAP-TTPKK-----PAPTT-----PKE-----PAPTTTKEPT 417
DB 894 VTSPSPQKTPKSVSLKGPAMTSKKATEIAASKDVSPSQFPKEVPLQLQHVPTSPPKSPV 953
QY 418 PTT-----PKPAPTTKEPAPTTTKE-PAPTAPKKPAPTTTKEPAP-----TTPKEP 463
DB 954 SDTSLGALTSPPPKGP-PATLAETPTYPKSKPAASKK-----TPATPSPEGVTAVPLEI 1008
QY 464 APTTTKEPSPTTKEPAPT-TTKSAPTT-TKE-----PAPTTTKGAPTT 505
DB 1009 PCKSKAPKATAPKESASISSKGRAPKTAVSKEIIPSGVTAVPLEISLPLIKETSKGA--T 1066
QY 506 PKEPSPTTK-BPAPTTTKEPAP-----TTPKKPAP--TTPKE-----PA 542
DB 1067 PGKSSASSPKRSPKTAGPKETPEGGVTAVPPEISLPPKETPONATPNESLAASSQKSPK 1126
QY 543 PTTTKEPAP---TTTKKPAPTAP-KEPAPTTPKET-----APTTPKKLTPTT 586
DB 1127 TSVPEKETPPGGVTAMPLEIPSAQKAPKTAVPKQIPTPEDAVILAGSPLSPKKASATA 1186
QY 587 EKLAPTTPE-----KPAP-----TTPPEELAPTT----- 609
DB 1187 PKRAPATPSVGVIAVSSEISPSPKTSKTAAPKENSATLPPKRSKTAAPKETPATSEB 1246
QY 610 -----PEEPTPTTPEEPAPTTKAA-----APNTPKEPAPTTPKPAPTT--K 651
DB 1247 VTAVPSEISPSPTTASKGVFVTLTPKGAENALAESPASPKVKPKTAAPSETSTTPSPQK 1306
QY 652 EPAPTTTKEAPT-----TPKGAPTTLKBP-----PTTPKKPAPKELAPTTTK 696
DB 1307 IPKVAGKEASATPPSKKTKTAVPKETSAPSEGVAVPLEIPEPSPK-APKTAAPKETP 1365
QY 697 EPT---STTSDKPAPTTT-PKT-----APTTP----- 719
DB 1366 APSEGATTAPVQIAPPSPRKGSKAGSKETPTTSPSEGVTAAPLEIPISSKTSKMASPK 1425
QY 720 -----KEPAPTTTKE-PAPTTPKGTAPT 741
DB 1426 ETLVTPSSKKLSQTVGPKET-SLEGATAVPLEIPSHKKAPKTVDPKQVPLTPSPK-DAPT 1484
QY 742 TLKEPAPTTPKKPKAPKELAPTTTKGPTSTTSGDKAPTTPKETAPT--PKPAPTTPKKA 800
DB 1485 TLAE-SPSSPKK-APKTAAPPSER-VTVVPPEKA--TPKASGTTASKVPVPAETQVA 1539
QY 801 -----PTTPTETPTTSEVSTPTTKE-----PTTIHKSPPDESTELSAPETPK 844
DB 1540 VSSRETPTVPAVPP-----VKNPSSHKTSKTIELKEAPATLPPSPKSPKIPSSKKAPR 1594
QY 845 ALENSPKP-PGVPTTKTPAATKPEMTTAKOTKTTERDLRTTPETT----- 888
DB 1595 T--GAPKEFPASPIK-----VTTSLAQATAPPSSLOKAPSTTIPKENLAAPVLPVS 1644
QY 889 -----TAAPKMTKETATTTTEKTETTESKITATTTOVTSTTTQDTTTPFKIT 931
DB 1645 SKSPAAPARASASLSPAAPAPOTAPKEATTIPSCKKAAT-----ETPIETS 1691
QY 932 TLKTTTLLAPKVVTTTKTTTITTEIMNKPEETAKPKDRATNSKAT----- 974
DB 1692 TAPSLGAPK--ETSETSVSKVLMSGSPPKKASSKRASTLPTATLPLSLKEASVLSFTATS 1749
QY 975 -----TPKPKQPTKAPK----- 987
DB 1750 SGKDSHISGVSDACSTGTTTPOASEKLPSKGPFTAEMLAAPAPESALAITAPIOKSPG 1809
QY 988 -----PTSTKKPKTMR--VRKKPTTPTPKMTSTMPENLNTSRIAEAMLOT 1032
DB 1810 ANSNSASSPKCPDPSSKDDTKGLPSAVALAPQTVPEK-----DTSKAETILLVS 1859

QY 1033 TTRPNQ-----TENSKL 1044
 Db 1860 PAKGSDCLHSPKPGVGSQVATPLAFTSDKVPPEAVSASVAPKAPAPASLTLAPGFPVAPL 1919
 QY 1045 VEVNPKSDAGAEGETHMLLRPHVHVEVTPDMYLPVRPNQGIINPMLSDETNICN 1104
 Db 1920 PPKQELLESAPGSVLESFSL-----PVAERDEL-----FLIPPEA-VSG 1960
 QY 1105 GKPDVGLTLRNGTLVAFRGHVFWLSPFSP-----PSPARRITEVWGIPSPDITVFT 1157
 Db 1961 GE-----PFQPILVNMPAPKEA-----GTPAPAPSA-- 1986
 QY 1158 RNCCEGTFPPKDSQYWRFTNDIKDAGYPKIFKFGG-----KQPVKNKSGSTESDSVPLEEGDSTQT 2018
 Db 1987 -----KQPVKNKSGSTESDSVPLEEGDSTQT 2018
 QY 1196 --LTGQIVAAALSTAKYKNWPESVYFFKRGSGIQYIYKQBPVQKCPGRRPALNYPVYGEM 1253
 Db 2019 ATQQAQIAAAAE-----IDEEPVSRAKQSR----- 2043
 QY 1254 TQVRRRPERAIGPSQTH-----TIRIQSPARLAYODKGVLHNEVKVSIILWRGLPNVVT 1308
 Db 2044 SEKKARKAMSKGLRQVTGVTIR-----KSKNLF-----VIT 2079
 QY 1309 SAISLPNIRKPDGY-----DYAFPSKQYINIDVPSRTARAITTR-SGOTLSKVWYN 1359
 Db 2080 -----KPDVYKSPASDTYIVFGEAKIEDLSQQAQIAAAAEKFKVQGEAVSNIQEN 2128

RESULT 5
 T45025
 mucin MUC5B, tracheobronchial [imported] - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T45025
 R:Desseyn, J.L.: Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 J. Biol. Chem. 272, 3168-3178, 1997
 A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
 A:Reference number: Z22899; MUID:97166151; PMID:9013550
 A:Accession: T45025
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-3570 <DES>
 A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
 A:Experimental source: placenta
 C:Genetics:
 A:Gene: MUC5B

Query Match 14.3%; Score 1048.4; DB 2; Length 3570;
 Best Local Similarity 13.5%; Pred. No. 1.4e-13;
 Matches 471; Conservative 159; Mismatches 545; Indels 2326; Gaps 107;

QY 23 SSQDLSSACAGCG-----EGYS-----RDATC----- 44
 Db 172 SAPGTTTCQPCQWTEWDEDPYKGEQLGGDVESYDKIRAAAGHLCQOPKIECOAESPP 231
 QY 45 -----NCD-----YNCQHYMECCPD----- 59
 Db 232 NWTLAQVGQKHCDVHFLGVCNWEQEGVFKMCMYNRVLVLCSSDHCRGRATTPPPTE 291
 QY 60 ----- 59
 Db 292 LETATTTTQALFSTPQPTSSPGLTRAPPASTAVPTLSEGLTSPTYSTLGTATGGR 351
 QY 60 ----- 59
 Db 352 QSAGSTEVPQVATSTLPTRSALPGTTGSLGTRWPSQPPLIAPTMTATSRAPGTAST 411
 QY 60 -----FKRVCIAELS-----CKGR----- 74
 Db 412 ASKEPLTSLAPTLSLSELSTQAEISTPRTETMTSPLNTTTSQGTTRCQPKCEWTEWFD 471

QY 75 -----FESFER-----GREC-----DQDAQ----- 89
 Db 472 VDFPTSGVASGDMETFERAAGGKMCWAPKIECRAENYPEVSDIQVQVLTCSLETGL 531
 QY 90 -CKKYDK-----CCPDYESFC-----AEVHNPTSP----- 114
 Db 532 TCKNEDQGRNMCNPNVNRVLCCDDY-SHCPSTLATSSTATPSSTPGTWTILTKPTTA 590
 QY 115 -----SSKAPP----- 121
 Db 591 TTTASTGSTATASSTQATAGTHVSTTATTPTVSSKATPFSSPGTATATPALRSTATTP 650
 QY 122 -----PSGASQTI----- 129
 Db 651 TATFTAIPTSSSLGTTWTRLSQLTTTMTATMTATPSSTPEIVHTSTVLTTATTGATGS 710
 QY 130 ----- 129
 Db 711 VATPSTPGTAHTTKVLTITTTTGTATPSSSPGRTLPVWISTTTTPTTRGSTVTPSSI 770
 QY 130 -----KSTTKSPKPPNKKTKKVKIESEIEITEVKDNKNR 164
 Db 771 PGTTHPTVLTTTITVATGSMATPSSSTQTSPTPSLTTT-----ATTITAGSTTNPS 825
 QY 165 TKKPTPKPPV----- 176
 Db 826 STPGTTPIPPVLTTTATTAAATSTVTPSSALGTHHTPPVVENTTATTGRLSPSSPHV 885
 QY 177 ---DEAGSLDNDGDFKVTTPD-----TSTOHNKVSSTPKI-----TTAKP 214
 Db 886 CTAWTSATSGI-LGTHITFPGTSHTPAATTGTTQ-----STPALSSPHSPSRITESP 940
 QY 215 INPRSLPNSDTSKETSITVNKE-----TTV----- 241
 Db 941 PSPGTTTPTGHTTATSRTATATSKRTSTLLPSQTSAPITTVVWGCEPQCAWSEWLD 1000
 QY 242 ----- 241
 Db 1001 YSYPMPGSGGDDFTYSNIRAGGAVCEQPLGLECRAQAQPGVPLRELGOVVECSLDLGL 1060
 QY 242 -----EYKETTITNKQSTDG-----KEKTTISAKE 266
 Db 1061 VCRNRQVQKFKMCFNVEIRVFCNMGHCPSTPATSTATPSSTPGTWTILTELTATT 1120
 QY 267 TQSIKTSKADLAPTSKVLAKTPKRAETTKG-----PAL----- 301
 Db 1121 TESTGSTA---TPTSTLRTAPPKVLITATTPTVSSKATPSSSPGTATAPALRSTA 1176
 QY 302 TTP-----KEPTPT-----TPK----- 313
 Db 1177 TPTATSVIPIPSSSLGTTWRLSQLTTTPTATMTSTATPSSTPETAHTSVLTATATTGA 1236
 QY 314 -----EPAST-----TPKEPTPTI-----KSAPTTPEP-----APT----- 342
 Db 1237 TGSVATPSTPGTAHTTKVPTTTTGTATPSSSPGTALTPPVWISTTTTPTTRGSTVTP 1296
 QY 343 -----TKSAPTTPK----- 351
 Db 1297 SSIPGTHTATVLTITTTTVAIGSMATPSSSTQTSPTPSLTTTATTATGSTTNPSST 1356
 QY 352 ---BPAP-----TTPKEPAPT-----TPKEPAPTTTKEPAPTTTKS-----APT----- 389
 Db 1357 PGTPIPPVLTATTATPAATSTVTPSSALGTTHTPPVNTTATTHGRLSPSSPHVTRT 1416
 QY 390 -----KBPAPTTPKKAPTT-----PKBPAPTTKPTT-----TT 420
 Db 1417 AWTGATSGTLGTHITBESTGTSHTPAATTGTTQHTSPALSSPHSPSRITESPSTGTT 1476
 QY 421 PKBPAPTTKEPAPTTPK----- 440
 Db 1477 PGHTTATSRITATATPKTRITSLPSSPTSAPITTVVWGCEPQCAWSEWLDYSYMPFG 1536
 QY 441 PTA-----PKP----- 447

| | | | | | | | |
|----|------|--|------|----|------|---|------|
| Db | 1537 | PSGDEDTYSNIRAGGAVCEQPLGLECRAQAGVRELQGVVECSLDGLVCRNREQ | 1596 | Db | 2617 | HTPPVPNTTATTHGRSLSPSPHVRVTAWTGATSGTLGTHTEPSTGSHTPAATTGTT | 2676 |
| Qy | 448 | -----APTPKEPAPTPKE----- | 462 | Qy | 743 | -----LKEPAP-----TTPKAPKELAPTTTKGPTSTISDKPAPTPPKETAPT--PKBP | 791 |
| Db | 1597 | VGKFMCFNYEIRVFCNYGHCPSPTATSSSTATPSSPTGTTWILTEQTAATTATTGST | 1656 | Db | 2677 | TTSFALSSPHSPSSRTTESPPSP-----GTTTPGHTTATSRITATATPSKTRTSTLLPSQP | 2732 |
| Qy | 463 | -----PAPTTPKEPSFTPK----- | 477 | Qy | 792 | ----- | 791 |
| Db | 1657 | ALPSSPTGTPPKVLTQSQTPTATSSKATSSSPRTATTLPLVLTSTATKSTATSFTPI | 1716 | Db | 2733 | TSAPITVVTTGCEPQCAWSEWLDYSYMPGPGSGDFDTYSNIRAGGAVCEQPLGLECR | 2792 |
| Qy | 478 | ----- | 477 | Qy | 792 | -----APTPPKP | 799 |
| Db | 1717 | PSSTLGTGTSQNRPHPMATMTIHPSSPTETHSTVLTATKATTRAITSMSPTSP | 1776 | Db | 2793 | ATAQCVPLGELQGVVECSLDGLVCRNREQVKFMCFNYEIRVFCNYGHCPSPTATS | 2852 |
| Qy | 478 | -----EPAPTTPKSAPT----- | 489 | Qy | 800 | APTTTETPTPTT-----SEVSTPTTTKEPTTHKSPDESTPELSAETPKALENSPKBPQVP | 856 |
| Db | 1777 | GTTWILTELTAAATTAALPHGTPSPSTPGTTWILTEPSTATVTVTGSTATASSFRATA | 1836 | Db | 2853 | STAMPSSTPGTTWILTELTATTTASTGSTATP--SSTP--GTAPPPKVLTSPTATPTAT | 2909 |
| Qy | 490 | ----- | 489 | Qy | 857 | TTKT-----PAATKPEMTTAKDKT-----TERDLRTTPTT--TAA | 891 |
| Db | 1837 | GILKVLSTATTPTVISSRATPSSSPGTATAPALRSTATTPTATSVTAIPSSSLGTAWT | 1896 | Db | 2910 | SSKATSSSPSPRTATTLPLVLTSTATKSTATSVTPIPSSTLGTTGLPEQTTTPVATMSTIH | 2969 |
| Qy | 490 | -----TTKEPAP | 496 | Qy | 892 | PKMTKETA--TTTEKTTESKITATTTQVTSNTTQDTP-----FKITTLTKTTT | 937 |
| Db | 1897 | RLSQTTTPTATMSTATPSSPTETVHTSTVLTATTATTRTGSVATPSSPTGTAHTTKVPTT | 1956 | Db | 2970 | PSSTPETHSTVLTATK-----ATTRATSSSTPSSPTGTTWILTELTATTTAGTGPTA | 3025 |
| Qy | 497 | TTT-----KSAPTTPKEP-----SPTT----- | 513 | Qy | 938 | -----LAPKVTTTKKTTT-----TTEIMNKPEE-----TAKPKDRAT-- | 969 |
| Db | 1957 | TTTGFTATPSSPGTALTPPVWISTTTPTTRGTSVTPSSIPGTHATVLTITTTTAT | 2016 | Db | 3026 | TPSSTPGTTWILTELTATTTASTGSTATLSSTPGTTWILTEPSTATVTAPPGSTAT | 3085 |
| Qy | 514 | -----TKEPAPTPKEPAP-----TTPKAPAP----- | 536 | Qy | 970 | NSKATTPKEP-----KPTKAPKPTSTKKPYT-----MPVRKPKTTPTRKMT | 1013 |
| Db | 2017 | GSMATPSSSTQTSPTPSLTTATTATGTTNPSSTPGTTPPVLTTTATTPAATSS | 2076 | Db | 3086 | SSTQATAGTHVSTTATTPTVTSKATPSSPGTATAPALRSTATTPTATSTFATPSSS | 3145 |
| Qy | 537 | TPKEPAPTPKEPAPTTT----- | 554 | Qy | 1014 | -----STWPELNPTSRIAEAMLOTT-----TRPNOTPNSKL | 1044 |
| Db | 2077 | TVTPSSALGTHTPPVVENTTATTHGRSLPSSPHVTVTAWTSAGILGTHITEPSTGT | 2136 | Db | 3146 | LGTWTRLSQTTPTATMSTATPSSPTETVHTSTVLTATTTATGATSVATPSSPTGTAH | 3205 |
| Qy | 555 | -----KKPAPT-----APKEPAPTP-----KETAPTTPKK----- | 580 | Qy | 1045 | VEVNPKSEDAG--GAEGETPHMLLRPHVFM-----BEVTPDMOYLP | 1083 |
| Db | 2137 | SHTPAATGTTQTPALSSPHSPSRTESPPSGTTTPGHTGTSRTATATPSKURTS | 2196 | Db | 3206 | TTKVPTTTTGTATPSSSPGTALTPPVWISTTTPTTTTSGSTVTPSSIPTGTHTA | 3265 |
| Qy | 581 | LTPTTPEKL----- | 589 | Qy | 1084 | RVEN-----QGIINPMLSDETNICNGKPVDLTLRN-----GTLVAF | 1122 |
| Db | 2197 | TLLPSSPTAPITTTVTTGCEPQCAWSEWLDYSYMPGPGSGDFDTYSNIRAGGAVCEQ | 2256 | Db | 3266 | RVLTTTATGSMATPSSSTQ-----SGTTPSLTTTATTTATGTTNPSSTPGT | 3318 |
| Qy | 590 | -----A | 590 | Qy | 1123 | RGHVFWMLSPSP----- | 1138 |
| Db | 2257 | PLGLECRAQAQGVPLRELQGVVECSLDGLVCRNREQVKFMCFNYEIRVFCNYGHC | 2316 | Db | 3319 | -----TPIPPVLTSMAATPAAATSSKATSSSPRTATTLPLVLTSTATKSTATSFTPI | 3370 |
| Qy | 591 | PTTPEKAPAPTPPEELAPTT-----PEEPTP-----TTPPEE | 620 | Qy | 1139 | ARRITEVWGIP-----SPIDTVFTRCNCE-----G | 1163 |
| Db | 2317 | PSTATSTATPSSPTGTTWILTKLTATTTTESTGSTATPSSSTQGPAGTHVSTTAT | 2376 | Db | 3371 | SSTLWTTWVPAQTTPMTSTMTSTHTSTPTTHSTVLTATTTATMTATNSTATPSSITLG | 3430 |
| Qy | 621 | PAPTPPKA----- | 628 | Qy | 1164 | KTFPKDSQYWRFTNDIKDAGYKPEIFKGFGLGCIIVAAL-----STAKYKNW | 1212 |
| Db | 2377 | PTVTSKATPSSPGTATAPALRSTATTPTATSTFATPSSSLGTTWTRLSQTTTPMATM | 2436 | Db | 3431 | TT-----RILTELTATTTAATGSTATLSST | 3457 |
| Qy | 629 | -----AAPNTPEKAPT-----TPKEPAPT--TPKEPAPTTPKET--APT | 665 | Qy | 1213 | PESVYFFKRGSGIQOYLYKQEPVOKPCGRPALNYPVYGMTQVRRRRFERAIGPSQHT | 1272 |
| Db | 2437 | STATPSSPTETVHTSTVLTATTATGATGSAVATPSSPTGTAHTTKVPTTTTGTFTVTPSS | 2496 | Db | 3458 | PGTTWI-----LTE-----PSTIAT | 3472 |
| Qy | 666 | PKGTA-----PTT-----LKBAPPTP | 682 | Qy | 1273 | IRIQVSPARLAYQDKGVILHNEVKVSIILWRGLPNVVTSAISLDPNIRKPDGYDYAFSKQY | 1332 |
| Db | 2497 | SFGTARTPPVWISTTTTPTSSSTVTPSSIPGTHPTVLTITTTQPVATGSMATPSSSTQ | 2556 | Db | 3473 | VWVPTGSTATSSLTGTAHT-----PKVVTAMATMP----- | 3503 |
| Qy | 683 | KKAPKELAPT-----TTKEPTSTSDKEP-----TTPKGTAPT--TPKEPAPT | 726 | Qy | 1333 | YNIDVPSRTARAITTRSGOTL | 1353 |
| Db | 2557 | TSGTTPSLTTATTATGATGSTATNPSSTPGTTPPELTATTATPAATSTVTPSSALGTT | 2616 | Db | 3504 | -----TATASTVPSSTV | 3516 |
| Qy | 727 | PKEPAPT-----TPKGPAPT | 742 | | | | |

cyst germination specific acidic repeat protein precursor - Phytophthora infestans
C;Species: Phytophthora infestans (potato late blight agent)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T31108
R;Goernhardt, B.

submitted to the EMBL Data Library, April 1998

A;Reference number: Z20986

A;Accession: T31108

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1489 <GO>

A;Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC7

C;Genetics:

A;Gene: car90

```
Query Match      14.0%; Score 1027.3; DB 2; Length 1489;
Best Local Similarity 24.2%; Pred. No. 6e-14;
Matches 430; Conservative 67; Mismatches 452; Indels 827; Gaps 84;

QY 41 DATCNDYNCOHYMEC-----CPDFKRYCTAELSCKGRCFESFGRGRCDCDAQ 89
DQ 32 DATYLSSESG--IPCAGVGAEPVGTACPKAGDVAT-----DCOPY 71
QY 90 CKYD--KCPDYBSPCAEVHN-----PTSPSSKKA-----119
DQ 72 LLSYNGAVCVAPVDAECALIHDDMWGCEFFPKTGYTSAVEAEITIAAYNGESSGWTGHDV 131
QY 120 -----IPCSGVGAEPVGTACPKAGDVAT-----DCOPY 71
DQ 132 VQVDEEEETPARVNVYDVTVDTPGVNCEVATEATQGHATEGGKYDTPSTGTQTDYG 191
QY 125 -----ASQTIKSTTKRSPKPPNKKTKKVIIESEB-----ITBVKNKKNRTKKTPK 172
DQ 192 NTHYGSTTTEGVTKGCGTGD-----AKVIDGETYLDYPTGITEIID-----235
QY 173 PPVVDEAGSLDNGDFKVVTPDTSST-----QHNKYSTSPKI- 209
DQ 236 -----GTTPGYGTITDGGTTGGYTTVDNTHETTBGAGYDAGTREETTPTVG 286
QY 210 -----TTAKPIN-----PPSLP-----222
DQ 287 YSTEETEGHVHVGYPSEDEATEPTGTYVPREETTAAPSEDTTVAPREVTVYAPTEKP 346
QY 223 -----PNSDT-----227
DQ 347 YDVEETTYVTEESTYAPTKSETNAPTERMHALEKPCDEVTVYAPTEETTVYAPTEET 406
QY 228 ---SKETSLTVNKETVE-TKET-TTNKQTSIDGKEKITSAKETOSIEKTSADLAPTS 282
DQ 407 YAPTEETTVAPTEETPEYPTTEETTYPTTEETTVAPTEET-----YAPTE 451
QY 283 KVLAKPTPKAEITTKGPAITPKETPTT-----PKE- 314
DQ 452 KTYAAT-----EETTVAPTEETPEYPTTEETTVAPTEETTVYAPTEETTVAPTEE 508
QY 315 -----PASTTPKEPTPTTIKSAPTPKPEAPT-TTKGAPTPKPEAPT--TTKEPAPTPK 367
DQ 509 TVYAPAEETPYPTTEET-TVAPTEETTVAPTEETTVYAPTEETTVYAPAEETPY 567
QY 368 EPAPITTKGPAITTKSAPT-----TPKEPAPTPKPEAPTTPKPEAPTTPKEPT 417
DQ 568 EPTTEETTVAPTEETTVYAPTEETTVYAPTEETTVYAPAEETPYPTTEETTVYAPT 626
QY 418 PTPPKPEP-----AP--TTKEPAPTPKPE-----APTAPKPEAPTTPKE 454
DQ 627 EETTVAPTEETTVYAPTEETTVYAPAEETPYPTTEETTVYAPTEETTVYAPTEET 686
QY 455 PAPTTPKPEAPT--TTKEPSPTTPKEP-----APT-TTKSAPT-----TTKEP- 494
DQ 687 YAPTEETTVYAPTEETTVYAPAEETPYPTTEETTVYAPTEETTVYAPTEETTVYAPTEET 746
QY 495 -----APT-TTKSAPT--TPKESPTTPKEP-----APTTPKEPAPTPKPEAPTTPKE 540
```

RESULT 7

T16251

hypotheetical protein F35A5.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C;Accession: T16251

```
Db 747 EETTVAPTEATTVAPTEETPVAPTEETTYEPTGETTVAPTEETTVAPTEETTVAPTEETT 806
QY 541 PAPT--TPKEPAPTTTKKPAPTAPKEP-----APTPKETAETTPPKKLPT 584
DQ 807 YAPTEETPYEPTTEETTVAPTEETPYEPTTEETTYPTTEETTVAPTEETTVAPTEETTVAPT 866
QY 585 TPEKLAPT--TPKEP-----APTPPELAPT-----TP 610
DQ 867 EETTVAPTEETPYEPTTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPAEETP 926
QY 611 EEPPTTPKEP-----APTPPKAAAPNTPKPEAPT--TPKEPAPTPKPEP-----APT 656
DQ 927 YEPTTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPAEETTVAPTEETTVAPT 986
QY 657 TPKEAPATTPKGTAP-----TTLKPEPAPTPPKKP-----APKE--LAPT-----693
DQ 987 EETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETT 1046
QY 694 -----TTKEPTSTTSKPA-----PTTPKGTAPATTPKGPAPTPPKPEAPTPPKGTAPT 741
DQ 1047 YASTEETTVAPTEETTVAPAEETPYEPTTEETTVAPTEETTVAPTEETTVAPTEETTVAPT 1106
QY 742 --TLKPEAPTPPKP-----APKE--LAPT-----TTKGPTSTTSKPAETP 780
DQ 1107 EETTVAPAEETPYEPTTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEET 1164
QY 781 KEIAPT--TPKEPAPTPPKP-----APTPPEPTTSEVSTPT--TTKEPTTIHKSP 830
DQ 1165 TVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPT--EETP 1222
QY 831 DEETPELSAETPKALENSPKRGVPTTKPAATKPEMTTAKDKTTERDLRTPETTTA 890
DQ 1223 YEPTTEETTVAPT--EETTVAPTEETTVAPT-----TEETTVAPTEETTVAPT 1263
QY 891 APKMTKET-----ATT--TEKTESKITATTQVTSIT--TQDIT--PFKLTIT--932
DQ 1264 AP--TEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPTEETTVAPT 1321
QY 933 LKTTTLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKPTSTK 992
DQ 1322 TEETTVAPTEETTVAPM-----EET-----PYEPAEESTSVSTE 1356
QY 993 KP-KTPRVRKPKKTTTPRKNMTMPDELNPTSIABAMLGQTTTPRPNQTPNSKLIVENPKS 1051
DQ 1357 KPCTBEFTDEBDEPTDE-----PSDEPTD-----EPTDEPTDLPTDE-----1395
QY 1052 EDAGGAEGEPTPHLLRPHVFMPEVTPDMDYLPKRPV--NOGIIINPMLSDETNICNGKPD 1109
DQ 1396 -----PSTPCDNQGI-----NGIGVE 1411
QY 1110 GLTTLRNGTLVAPRGHYFWMLSPFPSPARRITEVWGI-----PSPIDTVFTRCNCEGKT 1165
DQ 1412 NKVRYNNA-----GIYNTTGP-----1428
QY 1166 FPKDQYWRFT-----NDIKDAGYKPKFEGPG--GLTGQIVAAALSTA-----KYKNW 1212
DQ 1429 ---RNSQSHWSCRCSCYND-----PICHAFSPHQTSDSYCELTSTTSDEEDQONW 1477
QY 1213 PESVYFFKRGSGTQQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRRERFALGPSQHT 1272
DQ 1478 -----LAGNMDR-----1484
QY 1273 IRIQSPARLAYQDKGVLHNEVKVSIILMRGLPNVVVT 1308
DQ 1485 -----NVVT 1488
```

R,Leimbach, D.
submitted to the EMBL Data Library, January 1996
A;Description: The sequence of C. elegans cosmid F35A5.
A;Reference number: Z18485
A;Accession: T16251
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1274 <LEI>
A;Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB524
A;Experimental source: strain Bristol N2; clone F35A5
C;Genetics:
A;Gene: CESP:F35A5.1
A;Map position: X
A;Introns: 1272/2

Query Match 13.9%; Score 1018.4; DB 2; Length 1274;
Best Local Similarity 23.1%; Pred. No. 6.6e-14;
Matches 350; Conservative 106; Mismatches 346; Indels 716; Gaps 80;

```
QY 135 RSPKPPNKKTKK-----VIESEITEV-----KNNKNTKKKPTKPPVVDAG 180
D 135 RSPKPPNKKTKK-----VIESEITEV-----KNNKNTKKKPTKPPVVDAG 180
D 3 RAPPTPIKNPAKKWKPPWESVDEEEVEDEETPAPSKLEKKPSLKRKRDAPTREV----- 57
QY 181 SGLDNGDFKVTPTDSTTQHNKVSSTPKITITAKPI-NPRPSLPPNSDTSKETSLSVNKET 239
D 58 -----PSGAPSPVPIKNP----- 71
QY 240 TVETKETTTNNKQSTGDKGKTSKAKTQSIKTS-----KDLAPTSSKVLAK-----P 288
D 72 -----VKKWKAPWEDDEPMEEAAPAPVPAKVRDP 101
QY 289 TPRAETTKGPAITTPKEPTP-----TTPKEPASTTPKEPTP----- 326
D 102 SPK-----KVPA-----KPRDASGPKIWAAKKBETLPAVPPPTPVKNPVKKFAPWEDDEV 154
QY 327 -TKSAPTTP-----KEPA-----PTTKSAPTTPKEP-----APTTPKEPA----- 362
D 155 EDVKDAPTVAKTPVLKKKEPAAAKPRDPSPKKAPSKHEHDPIVPTPIKNPAKKWK 214
QY 363 -----PTTP-KEPAPTTPKAPTTPKAPTTP-----KEPAPT-----TPKKPAPTTPKEP 408
D 215 PWEDDEVPTBEIKEPEPATRKVPA--LKKKEPSTSVKPSVSDPSPTKKVFKKEPEVP--- 269
QY 409 APITTPKEPT-----PTTPKEPAPT-----TKEPAPT-----PKEPAPT-----APK 445
D 270 -PTIKNPTKKKPPWEDDEVVEEVKEPPVPEKKAPVLKKDPAKAPAAKARDPSKAPK 328
QY 446 KPAPTTP-----KEPA-----PTTP-----KEPAPTTPKE 470
D 329 KVEPSSPVVPTPVKNPVKKYPWVEVDEFAEEVKKPSAPEKKTPLVKRKEPSSSTP 388
QY 471 PSPTTPKEPAP-----TTTKSA-----PTTPKEPA-----PTT 498
D 389 SSDSPKKAAPAVKPRDSSPKKATPLQADPKAQAEVPTPVKNPVKKYKPPWEDDEV 448
QY 499 TKSAPTTP-----KEPSTTPKAPT-----TPKEPAPTTPKAPTTPKE----- 540
D 449 EVKQPEAPAKTPVLKKEPAKOTAKPAKTSKTETPEKDPVKPRDSSPKKVAAPKPSA 508
QY 541 PAPPTTP-KEPA-----PTTKKPAAPTAPKEPAPT----- 568
D 509 QAPATPVKNPVKKWRPPWEDDEVTPADDVSKPTDAKKTPLSKADPAKESLKPADTKA 568
QY 569 -----TPKETAPTTPKLTPT-----TPKLPAP-----TTPKEKAP 599
D 569 PAKPRDPSPKVAPTAPEKKTPLVLAKKEPAGPADSKTKEPKSKPRDPSPKKAVPAKVP 628
QY 600 TPELAPTPEEPTTPPEPAPTTPKAAAPNTKPEAPTTPKEP-----AP 647
D 629 KT--EVAAPAAVKKEPIS--KPKOTAPKKAEPNSPVVP-PTPVKNPVKKWPPWEDDDAP 683
QY 648 TTP-----KEPAPTTP--KETAPTTPKG-----TAPTTLKEPAP--TT 681
```

```
D 684 AKPVSLPEPEKKTPLVLAKKAPTDPSEAAADPVSGFSSKDKLAKKAPVAKPRDPSPMKAV 743
QY 682 PKKPAK-ELAPTTTK--EPTSTTSD-----KPAPTTPKGTAPTTP-KEP----- 722
D 744 PIKPAKTEVPPAVVKRPEVAKSRDPSPKKAAEPNSP--VVPPTPVKNPVKKWKPWE 801
QY 723 -----APTTPKEPA-----PTTPKGTAPTTLKEPAPTTP 750
D 802 DDDAPAEVNVPEPEKKTPLVLAKKTPVKPRDPSPKKAVPAKSPSTKTDAPVSVKKPEPVS 861
QY 751 -PKKPAKPE-----LAPTTTKGPT-----STTSKPAPTTPKETAFT-TPKE 790
D 862 KPKEPSPKKAEPNSPVVPTPVKNPVKKWPPWEDDEVTEEVVKPSSEPKKTPVLAKKE 921
QY 791 P-----AP-----TTPKAPAP--TTPETTP-----PTTSEVSTPTTTKE 822
D 922 PEKPAQAPKVAAPKPRDPSPKKAVPEKEPAKAAKPRDLSPKKALPIPAANTQEAFTPVKN 981
QY 823 PTTIHKSP--DESTPELSAEP-----TPKALENSPKPEGVPTTK--TPAATKPEMT 869
D 982 PVKKWPPWEDDDDEP--AEPVSAPEPEKKTPLVLAKKAPAKPRDPSPKAAPVAAPKPD-- 1036
QY 870 TTAADKKTTERDLRTTETTTAAAPKMTKE-----TATTTEKTTESKI 910
D 1037 -----PKIPEVPTPVKNPVKKWPPWEDDEVSEPEVSAPEPEK----- 1075
QY 911 TATTQVTSITTQDTPFKITTLTKTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATN 970
D 1076 -----KTPVLAKKAPTTPKATKPDSEAAADPVSGTSPKPKLS 1112
QY 971 SKATTPKPKPTKAPK---KPTSTKPKTTPRVRKPK-----TTPPTP 1009
D 1113 KKAPVEKP-KPTTDPKDDKLPSPAKPEKAPAPAAKWKPMVDDDPDEPEADFTVPAP 1171
QY 1010 RKMSTSM-----PELNPTSRIAEAMLQTTTTRNQNTNSKLVENVKPSSEDAGG 1056
D 1172 SKKPDTEBDPADLGGPKTKDKPLNKKA-----PAEKPTKPKPEVSKPEPK----- 1218
QY 1057 AEGETPHMLLRPHVFMPEVTPMDYLPRVPMNOGIIINPMLSDETNIENGKPVDDGLTLRN 1116
D 1219 -----PTEPKK-----PAAPKK-----WKPP----- 1234
QY 1117 GTLVAFRGHVFWMLSP-----FSPSPARRITEVMGIPSPIDIVTFRNCCEGKTFPPKD 1170
D 1235 -----WEDDPDEPEADFTMPAPKKPDTE-----DPADSL----- 1263
QY 1171 SQYWRFTNDIKDAGYKPK 1188
D 1264 -----GGPKP 1268
```

RESULT 8

T34513

hypotheical protein ZK783.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C;Accession: T34513

R;Favella, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A;Description: The sequence of C. elegans cosmid ZK783.

A;Reference number: Z21536

A;Accession: T34513

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-3507 <FAV>

A;Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:Z

A;Experimental source: strain Bristol N2; clone ZK783

C;Genetics:

A;Gene: CESP:ZK783.1

A;Map position: 3

A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 1175/3; 1207/1; 1409/2;

3504/1

| | | | |
|--|---|--|--------|
| Query Match 13.4%; Score 983.3; DB 2; Length 3507; Best Local Similarity 13.2%; Pred. No. 3.3e-12; Matches 472; Conservative 206; Mismatches 519; Indels 2377; Gaps 118; | | | |
| QY 7 | PIYLLLLSVFVIOQVS | QDL | SSC 30 |
| Db 32 | PTFVNFDTSTVICOHSSDPTDLHNMSSLCDGKQDCFNVPAMHDEVFPYCEHKQSTC | 91 | |
| QY 31 | AGR | -CGEGYSRDA | TNC |
| Db 92 | SGKAGLYDGAPOCYCDSGFSGACELQDKNECLEHPCHMMAQCQNTLGSYECRCCLPGY | 151 | |
| QY 47 | ---DYNCOHYMEC | ---CPDFKRV | 63 |
| Db 152 | EGNGHECTDIDEGSKLTSRCPESHKCNLPPTYNCYTCQFTPKGNQSGLDKCADINE | 211 | |
| QY 64 | ---CTAELSK | ---GRCFESFERGRE | 83 |
| Db 212 | CETGAHNCDADEICENSIGSKVCNKGSPGYELIDGKCEDVNECGSEKLHKCDVRADCVN | 271 | |
| QY 84 | --- | ---CDC | 86 |
| Db 272 | TIGGYECEBEGFEGDKNCQKCGHSVKLPKSSCRKNSAICDRHASCHIVLIDCDCKT | 331 | |
| QY 87 | --- | ---DAOCK | 92 |
| Db 332 | GYTGDGITCHDINECAKTPCSDGGRCLNLDGYYVCCNGQDDATCIKQGAFCSCGGC | 391 | |
| QY 93 | --- | ---YDKC | 112 |
| Db 392 | DNAICSNATCACIDGFRGDPHKKCVDINECVENDSVCGVGDRCVNLFGGFKCQHGSTE | 451 | |
| QY 113 | PPSSKKA | ---PPPSGASQTIKSTTKRSPK | 138 |
| Db 452 | AECTDQAFSSDSSTISHGADFTTTGEOIIEGSGSIQTSSGSLTV | 509 | |
| QY 139 | --- | ---PPN | 141 |
| Db 510 | TSGRLACTSYCPNSENCEVGVCEBVCYGGNVLVGCEDIDECITEICNIEANWCYNLIG | 569 | |
| QY 142 | --- | ---KKTKKVIESEE | 153 |
| Db 570 | GFVCCNPTNATHDDCIDFLTKVIYAMIIIFLLKGLITKKEGLHVIGNEEDTVVATR | 629 | |
| QY 154 | --- | ---ITEVKDNKN | 168 |
| Db 630 | SNHSTDQLITVQQSRNFSTQIILTRGVSSGEAVTQTDDAFGLEISAADLAGSG | 699 | |
| QY 169 | --- | ---PTPKPPVVDEAGSL | 187 |
| Db 690 | SGITLPTTLEPKIE | ---GSGKASGVWVTEDEGEDEDELMEEGSGSWSTTINGTIGTSR | 747 |
| QY 188 | --- | ---FKVTT | 198 |
| Db 748 | SEGTRVITTLGEDGEPEATATKPGISAPDKTGEKSTESDBEKLTVBKDGEAQS | 807 | |
| QY 199 | --- | --- | 198 |
| Db 808 | SATSSGKSEATSGSSSSAKSGTSGSEAGSSGASSSGSGVSGSGSVSTESGSGFTS | 867 | |
| QY 199 | --- | ---QHNKVSTSPKI | 212 |
| Db 868 | SSGSVGEATSGTVDSGSPKSSTEKLPFTKNKEKSPISGSDTTGKSSSEBTSR | 927 | |
| QY 213 | KPINRPSLPPNS | ---DTSKETSITVWK | 237 |
| Db 928 | KPIEGSDSITGSGSGGEWFETGSKGHFSGSKSVTSKGPTQSGAEGSGSGPKVPKPG | 987 | |
| QY 238 | --- | ---ETTVETKETTT | 270 |
| Db 988 | APETITDGESSSTGDKSGKPADKSNKNVPTGDKNPDITDDGEDSTSETSGEQG | 1047 | |
| QY 271 | EXTSAKDLP | --- | 289 |

| | | | |
|---------|---|-------------------------|-----|
| Db 1048 | PKGSKGQPPGDKGSEVKKPTSEVDGPGNLSGTGKGNVPLKPTDLPPEGSGILTTSSGG | 1107 | |
| QY 290 | --- | ---PK | 316 |
| Db 1108 | KNSTFEHGTKLRLPFPKTKEDKSSETPOLGLEISAKKPEPEDGTSKEVGLIEMWESTPG | 1167 | |
| QY 317 | STT | --- | 319 |
| Db 1168 | STTILDSVGLIEISGDLTKATKKPHVEIBSGTGDBEITATTTRDVSSTKKPRVEVDGG | 1227 | |
| QY 320 | --- | ---PKEPTTTIKSA | 338 |
| Db 1228 | DNGETSGVDKPTTAPTPSSAESSTSRPTTSEASPEGSGEAGVPESPDGSGESSTS | 1287 | |
| QY 339 | --- | ---APTTSKAPTTPKEPATT | 373 |
| Db 1288 | APDGVSTSSATAPEVTTASSTFPDAVERSGIPSTSKPTAEPLTT | 1343 | |
| QY 374 | --- | ---TKEPAPTTKAPTTPKEPATT | 412 |
| Db 1344 | EGSGTEBTLPTTEGGESTTSGAPTV | 1401 | |
| QY 413 | --- | --- | 415 |
| Db 1402 | PQANDSSVENTKCTSSDEGLDALCERRTGVCRCEPFGAGPPKSCVDVDECATGDHNC | 1461 | |
| QY 416 | --- | --- | 417 |
| Db 1462 | HESARCONVVGACPTGFRKADGSCQDIDBCTEHNSTCCGANAKCVNKPQTSYCEC | 1521 | |
| QY 418 | --- | --- | 429 |
| Db 1522 | ENGFLGDGYQCVPTT | 1580 | |
| QY 430 | --- | --- | 433 |
| Db 1581 | VCEDINECVAEKAPCSLNANCVNMNGTFCSCQKQGVRGDFMCTDINECDRHPCHPAE | 1640 | |
| QY 434 | --- | --- | 438 |
| Db 1641 | CTNLEGSFKCECHSGFEGDGIKCTNPLERSCEDVEKFCGRVDHVSCLSVRIYNGSLSV | 1700 | |
| QY 439 | --- | --- | 440 |
| Db 1701 | CECEPGRFPEKESNSCVDIDECESRNCDPASAVCVNTEGSYRCEAEGYEGGVCTD | 1760 | |
| QY 441 | --- | --- | 445 |
| Db 1761 | IDECDRMAGCDSMAMCINRMSCGCKCMAGYTGDGATCIKIBEEPKSDKTACTDEWSRL | 1820 | |
| QY 446 | --- | --- | 457 |
| Db 1821 | CELEKKQCTVDEEVQCGACLPGHHPINGTCQSLQISGLCAQKNDCKNHAECIDHPS | 1880 | |
| QY 458 | --- | --- | 458 |
| Db 1881 | HFCSOPDGFIDGMI CDDVDECNAGMCDDEKTCENTIGSFNCVCLGFKKVDKCVVD | 1940 | |
| QY 459 | TKPEP | --- | 506 |
| Db 1941 | EKKQNRKIEIDENSSSSNGQKPTTKGVTSSTATSESTTAPHVTTISSTT | 2000 | |
| QY 507 | KEPSTTTTKEPATTTPKEPATTTPKAPATTTPKEPATTTPKEPATTTPKAPAPKEPA | 566 | |
| Db 2001 | KDWTSSKSPENVTMESSEPEVSTSSKSTTASSETTVSSTPSESS | 2053 | |
| QY 567 | PTTPKETAPTTPKLTPTTPE | 611 | |
| Db 2054 | PATTTTEV | 2107 | |
| QY 612 | BPTPTTREPAPTTPKAAAPTTPKEP | 661 | |

Db 2108 SVTSTVPETSKSVLSSEAPVTSSTPEVHTSSSTKPSLSASSTTGDTNSTTSTSSLAS 2167
QY 662 --APTTPKGT--APTTLKEPAPTTPKKAPAKELAPT-----TTKEP 698
Db 2168 VKSATSAPGTSASVAPVKLSLSPDV-SQPSTKTFDATESSTVOAGSETSGTSVKSTSEP 2226
QY 699 -----TSTSDKAPATTGKTAPTTPKEPAPTTPKEPAPTTPKG-----737
Db 2227 ESHVTKLSITSSNPFSSVPVTSKSTFTVPEST--EQPTSTPSSGSLTPMNSSEVLTT 2284
QY 738 -----TAPTTLKEPAPT-TPKKAPAKEL-----APTTPKGT-----768
Db 2285 SEPHVLSLSPDVSSQSTTPNNLSSESTVETPKTSSEVSLNSEESPSTTEAPTLLSPDIL 2344
QY 769 STT-----SDKP-----APTTP-----780
Db 2345 STTTNNLSQSSTVSTEDRSBISSENSEKPTSAPELVTSSVTHVASSSPDPVPTESSEPDLL 2404
QY 781 -----KETAPTP-----KEPAPTTPKKP-----APTTP-----805
Db 2405 TGSSTENIPRASSKQTISSPTPTDTTASBEPTKSTMSMDLSTSNVLSSESSTIPSSS 2464
QY 806 -----TPPTTSEVSTPTTKEPT--825
Db 2465 KSPVSSSTEGISVVTSEFKVPPESTISSVLEEDLTKTTPSPILLEETTTASETSEPLETD 2524
QY 826 -----IHK--SPDESTPELSAEPPTKALENSP-KEPG-----VPTT-----858
Db 2525 SLTVSVRIHELTTSENVPKSESTTSSSESKSPQBPAGILTSTVVVPTSSVSLITASE 2584
QY 859 -----KTPAATPEM-----TTTAKDKTTERDLRTTPETTTAAPKMTKE 897
Db 2585 IEAITSNTPPKQGTPTITPKSLVKSTTSPSTVTSSEPSSEKRTTSTVSTVSTTTTEE 2644
QY 898 TAIT-----TEKTTESKITATTQVTS-----TTTQDIT 926
Db 2645 TTTSESILLTAAPKPTSTETSEFAPTTPAKTSEKPSNVSSSRKSTENVEFTSQSG 2704
QY 927 PFKITLLKT-----TLAPKVTTKKTIITTT-----952
Db 2705 SLESSTMSSTSEPETNAPAVTVSSEASTTLENSSTSPSTSEASVKLSLFPESITS 2764
QY 953 -----EIMNKPE-----960
Db 2765 EAVTVSSRAPAEITMSSEHREISTVSSSESEPELIPLSTTVSPNVVTASSIPSEBILSS 2824
QY 961 -----TAKPKD-----RATNSKA-----TTPKP 978
Db 2825 VTSSSTPRVRLITGTPDDLLIVSVTVPSHGNRRQNTASSVPNSNSTPILPSESLLTTPQ 2884
QY 979 QKPKAKPKPTSTKKPKTMPRVKPK-----TTPTP-----RKMTSTWPEL 1019
Db 2885 PPTTTTAKPATTSKGRPPSIOPPAEMFTTTPAPPFPSPNGYGEBETNQEEQVST--2940
QY 1020 NPTSRIAEAMLQTTT-----1034
Db 2941 ---TTEAPSLCSTVTCHSLATCEOSTGVCIORDGFIGDTTACSKKSTADCSILPSLCA 2998
QY 1035 ---RPNQTPNSKLVEVNPKSEBAGAGETPHMLLRPHVFMPEVTPDMXYLPRVENQI 1091
Db 2999 KAKCDNSTRSCEC-----DAG-----YI-----GDGYV 3021
QY 1092 INPMLSD---ETNLCNGKPV-----DGLTLR-----N 1116
Db 3022 CSPHPQDCVLRDLNLCSPBAVCQNRRCOCLPGFTGDGVKCVSIHERASNCQDANAHCVG 3081
QY 1117 GTLVAFRGHYFWMLSPFPSPPARRITEVWG-----IPSPIDTV-FT-----1157
Db 3082 GTTCKCNPGYP-----GNGLCCVPDPLDCVHFTGICHPNVACNSES 3122
QY 1158 -RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFG-----ITGQIVAAALS 1205
Db 3123 RQCQCS-----SGFSGNGVSCFPKQKSCRTDKSVCA--3152

RESULT 9

T29757

protein UNC-89 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C:Accession: T29757

R:Du, Z.; Le, T.T.; Wilson, R.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid C09D1.

A:Reference number: Z20679

A:Accession: T29757

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-6642 <DUZ>

A:Cross-references: EMBL:AF003131; PDB:AA54132.1; GSPDB:GN00019; CESP:unc-89

A:Experimental source: strain Bristol N2; clone C09D1

C:Genetics:

A:Gene: CESP:unc-89

A:Map position: 1

A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1;

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 13.3%; Score 971.4; DB 2; Length 6642;

Best Local Similarity 12.8%; Pred. No. 2.3e-11;

Matches 502; Conservative 197; Mismatches 478; Indels 2756; Gaps 132;

QY 34 CGEGYSRDATCNC-----DYNQCH-----52

Db 659 CKEGDVDFCEVGEVGEPELVMLVDQQLRPSHDFRLQYDQGTAKLEIRDAQDDTGYY 718

QY 53 -----YMECCPD-----FKRVCTAB--67

Db 719 TVKIQNEFGSIESKAELFVQADPDKNHVAPEFQATIEVEC--DEGEVRFKSVITGDPN 776

QY 68 -----LSCKGRCFSEFERGECDC 86

Db 777 PEIIWFINGKPLSESEVKFISEDGICILITIKDVTRHFDGMVTCQGS-----NRLGSASC 831

QY 87 DAQCKKYDKCCPDYESCAEVHNTSPSSKKA-----PPPS--123

Db 832 DGRLLK-----VRVPAPPTFNKPLEDKTVQEKSTVFEVDVSGWPEPILT 876

QY 124 -----GAS-----126

Db 877 FTLCGKELKNGEVEIVGHDFYRISIPNTSMKDHDGEIVAKAQNHEHGTAESRLTVE 936

QY 127 -----QTIK-----STTKRSPKP-----139

```

Db      937 QEEERSAPFLKDIEDQTVKTEFAVFEITVRGNPNPEVTFWFINGHKMDQSGPVKIE 996
QY      140 -----
Db      997 AHNHDHKLITDSAOYAGTVLCRAENAVGRFETKARLVVLAPEKQKPKPFVEILLVDKTET 1056
QY      158 KDNKKNRTKKXPT-----PKPPV-----
Db      1057 VDN-----TVVFVRVEGEFKPTVTWYLGEBELKQSDRVEIREPDGSIKISIRNIK 1107
QY      176 VDEAG-----SGLDNGDFKVT-----
Db      1108 IEDAGEIRAVATNEGSDETKAKLTVOKKPFAPEFDLRFVSLTVEKGSFAVFAHFGIP 1167
QY      192 -----TPDTS-----TQHNKVS-----
Db      1168 LPTYVSVNGRKRVDGEGARVTRDESTVCGASILITIDTATYYSEVNHLLTISVAENTLG 1227
QY      205 -----TSPKLTITAKPINPRESLPNS 225
Db      1228 AEETGAQLITIEPKKESVVVEKQDLSSEVQXELIAQOVKEASPEATTT----- 1274
QY      226 DTSKETSUTVNKETVETKETT-----
Db      1275 -ITMETSLSITKTTMTMTTEVTSVGGVTVBTEKESSESATTVIGGGSGGVTEGSI VSK 1333
QY      250 -----NKQTSSTGCKEKTSAKETQS 269
Db      1334 IEVVKTSQDVRBEGTPKRRVSFAABELPKEVIDSDRKKKSPSPDKKE-----S 1385
QY      270 IEKTSADLAPTSKVLAKTPKAEITTTKGPALTTPKPTPTTP-----KEPASITTPKEPT- 324
Db      1386 PEKTEEPASPTKKT-----GEEVKSPEKSPASPTKKEKSPAABEEVKSPTK 1432
QY      325 -----PTTIKGAPTTP-----KEPAPTTTKSAPTTPKPEAPTTTKEPAPTTP----- 366
Db      1433 KEKSPSPPTKKEKSPSPSKTKTGDEVKEKSP-----PKSP-----TKKEKSPKEKPEDV 1480
QY      367 -----KEPAPTTT-----KEPAPTTTKSAPT-----KEPAPTTTKSAPT 387
Db      1481 KSPVKKEKSPDATNIVEVSETTIEKTETTMTTMTHESESRTSVKKEKTEPKVDKPK 1540
QY      388 TP-----KEPAP-TTPKPEAPTTKPEAPTTKPEPTTPKPE-PA 425
Db      1541 SPTKOKSPEKSIETEEKSPVKKEKSPKEVBEKPASTPKKEKSPKASPTKKSENVKS 1600
QY      426 PTTKEPAP-----TTPKEPAP-TAPKKPAPTTPKPEAPTTTPKPEAPTTTKEPSPTTP 476
Db      1601 PTKKEKSPKSVVEELKSPKEKSPKEKADDPKSPTKKE-----KSPEKSATEDVKSPTK 1654
QY      477 KEPAPTTTKSAPTTPKPEAPTTKSAPT-----TPKEPSPTTKEPAPTTTPKPEP 525
Db      1655 KEKSPKEVEKPT-----SPTKKESSPTKKTDDRVKSPTKKEKSPQVVEEKSPASPTKKEK 1709
QY      526 AP-----TTPKPEAPTTKPE-PAPTTTPKPEAP-----TTTKK--PAPTAPKE 564
Db      1710 SPEKSVVEEVKSPKESKAEKPKSPTKKEKSPKSAEEVKSPTKKEKSPKESKABEK 1769
QY      565 PAPTTPKEAPTTPKKL-----TPTTPEKLAPTTPKPEAPTTTPEELAP--TTPPEPTPT 616
Db      1770 PKSPTKKESPV---KQADDEVKSPTKKEKSPKEVBEKPASTPKKEKTPKESAABELKSP 1826
QY      617 TPEEPAPTTPKAAAPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKETAPTTPKPGAPTTLKE 676
Db      1827 TKKEKSPSSPTKKTGDESKEKSPKEPE---RPKSP---TPKSPPGSPK-----KKKS 1874
QY      677 PAPTTPKPKAPK-----ELA----- 691
Db      1875 KSPEAEKFPAPKLTRODLKLOTVNKTLAHEVVVVVEHATECKWFLDGKEITTAOGVTVSKD 1934
QY      692 -----PTTTKEPTST----- 701

```

```

Db      1935 DQFERCSDITTFMGSGTIVSVASNAAGSVETKTTELKVLETPKTKBEFTDKLDMEVT 1994
QY      702 -----
Db      1995 KGDTVQMDVIALHSPLYKQWQNGNLELDGKNGVITIKNEENKSSLIIPNAQDSGKITVEAS 2054
QY      702 ----TSUKPAPT--PKGTAPTTPKEPAPTTPKE-----PAPT----- 733
Db      2055 NEVGSSESSAQLITVNPSTTPIVVDGPKSVTIKETETAERKATISGGFPAPTVPKWITNEKI 2114
QY      734 -----TPK-- 736
Db      2115 VEESRTITTIKTEDVYTIKISNAKIEQTGVTKVTAQNSAGQDSQADLKVPENVKAPFK 2174
QY      737 -----GTAPTTKE 745
Db      2175 SQLTKVADEGEPLRWNLDELDPSPGTEVSWLLNGQPLTKSDTVQVVDHGDGTVHYVIAE 2234
QY      746 PAP-----TTPK----- 752
Db      2235 AKPEMSGTLTAKAANAAGBCEETSAKVTVNGNGNKKPEFVQAPQNHETTLEESVKFSAIVTG 2294
QY      753 KPAPK----- 757
Db      2295 KPENVTWYLNKKLIQSEEVKVKYVHETGKTSIRIQPLMEHNGTIRVEAENVSGKVOA 2354
QY      758 -----ELAPTTT-----KGPTST 770
Db      2355 TAQLKVDKKTVEPKFTTNDMDROVKEGEDVKFTANVGEYPEPSVAWTLNGEPVSKHPNIT 2414
QY      771 TSKP-----APTTPKET-----APTTPK----- 789
Db      2415 VTDKGEHTTIEISAVTPEQAGELSCAENPVGSKKRDVQLAVKVGADPTFAKNLEDELI 2474
QY      790 ----- 789
Db      2475 TEGELTLMDAKLNIIVKPKPILTWLKDQVEITSDGHYKIVEEDGSLKLSILOTKLEDKGR 2534
QY      790 -----EPAPT 795
Db      2535 ITIKABSEFGVAECASLGVVKGKPAQFQSDIAPINLTGDTLECKLLITGDPPTFV 2594
QY      796 -----PKK----- 798
Db      2595 KWIGTQVLCATEDTEISNANGVYTMKIHGVTADMTGKIKCVAINKAGEVSTEGPLKVA 2654
QY      799 PAPTTPET-----PPTTS-----EVSTPT 818
Db      2655 PIPVEFETSLCDATCREGDTLKLRAVLLGEPEPVVWYVNGKKLEESQNTIKHSEKGYT 2714
QY      819 TTKEPTT-----IHK 828
Db      2715 VTIKDITCDYSGQWCEAINEYKATSEATLLVLRGEPDPDFLEWLSNVBARGTGVVHK 2774
QY      829 -----SPD----- 831
Db      2775 VFTGDPKPSLWYINKEILLNSDLVTIIVDDKTSTLTINSFNPDPVHVGBIICAEADAG 2834
QY      832 -----ESTPELSAE-----PTPKALENSPK-- 851
Db      2835 EVSCTANMITYTSDMFSESESESAQAEFVGGDDLTEDESLEEMHRTPTPW-----APAFIT 2891
QY      852 -----EPGVPTTK----- 859
Db      2892 KIKDTKAKKXGSAVPECVVDPDTKGVCCKWLDKGKIELIARIVQTRTGPBGHITOELVL 2951
QY      860 -----TPAAT-----KPEMTTAKOKTTERDLRT 893
Db      2952 DNVTPEDAGKYTCIVENTAGKDTCEATLTVIESLEKKSEKKAPEFIVALQDKTKTKSEK 3011
QY      884 T-----PETTTAAPKMTKETATTTTEKT-TESKITATITQ-----VTSTTQDTTTF 928
Db      3012 VLECKVIGPKFVSWLHNVNREKNPNEKTIITQESITVESVEGVERVTITSSSE----- 3066

```

```
QY 929 KITT-----KTTTAPKVTTKT-----ITTTIMNKP-----958
Db 3067 -----LSHQKYTCIAENTEGTSKTEAPLTVQGEAPVFTKELONKELSIGKLVLSGVK 3122
QY 959 -----ESTAPKDR-----ATNS 971
Db 3123 SPOPHVDFYSFSETTKVETKITSSRIAIEHDQTNTHWRVISOITKEDIVSYKATINS 3182
QY 972 KAITPKPKPT---KAP-----KKPTSTK-----KP-----994
Db 3183 IGTATSTKITTKEAPVFGGLKTSVKEKEIKMEVKVGGSPADVVEWFKDKPVSEGD 3242
QY 995 -----KTMPR-----VRPKPTT-----PT-PRKM 1012
Db 3243 NHEMKKNPETGVFTLVVQKAATTDAGKYTAKASNPAGTAESSABAEVTSLEKPTFREL 3302
QY 1013 TSTMPELNPSTRIAEAMLQTTT-----PNOTENSKLV-----1045
Db 3303 VTTEVKINET-----ATLSVTVKGVPPSPVSEWLKOGQPVQTDSSHVIKAVEGSGSYSITI 3357
QY 1046 -----EVNPKSEDAGGAEGETPHMLLR---PHVFMPEVTP-----1077
Db 3358 KDARLEDSGKYACRATNP---AGEAKTEANFAVVKNLVPPFEFVEKLSPLEVKEKESTTL 3413
QY 1078 -----DMD-----YLPVPVNO 1088
Db 3414 SVKVVGTPSPVSEWFKDDTPISIDNVHVIOQTAVGFSFLTINDARQDVGIVSCRRNE 3473
QY 1089 -----GII-----INPM-----LSDETNIC 1103
Db 3474 AGEALTTANFGLIIRDSIPPEFTQKRLPLEVREBQETLDLKVTVIGTPVNPVWFKDD---3529
QY 1104 NGKEVD-----GLTTNGLTVAFRGH-----1125
Db 3530 --XPINIDNSHIPAKDESGHHTL--TIKQARGEDVGVTCKATNEAGEAKTTANMAVQ 3584
QY 1126 -----YFWMLSP---SPSPARRITVWGPSP---1152
Db 3585 ETEAPLFGVLQKPYEVEGQGPABLVRVVEGPEPEVKWFGDGVPIAIDNOHVIEKKGEN 3644
QY 1153 -----1152
Db 3645 GSHTLVIKDTNNADFGKYTCQATNKAGKDETVGBLKTPKPSFEKQTAEEVKPLIEPLKE 3704
QY 1153 -----DTVFRPCNCEGKTF---FFKDSQWRF-----1176
Db 3705 TFAVEGDTVVLECKVKNKESHPIKFFKNDQPVETGQHMQLVLEDGNIKLTIQNAKKEDV 3764
QY 1177 -----TN-DIK-----DAGY 1185
Db 3765 GAYCEAVNVAGKANTNADLKIQFAAKVEHVHVDSEGOLEBIGQFETVGDTSASKTTDGR 3824
QY 1186 PKPIF-----KGFGGLT 1197
Db 3825 GAPEFVELLRCTVTEKQQAILCKVKGEPRPKIKWTKEGVEMSARVRAEHKDDGILT 3884
QY 1198 -----GOIVAAIISTA-----1207
Db 3885 LTFDNTVQADAGEYRCEAEENYGSAAWTEGPIIVTLEGAPKIDGAPDFLPQVPKPAVTVTG 3944
QY 1208 -----KYKNWPE-SVYFPRKGGIO-----OY---1228
Db 3945 ETAVLEGKISGPKSPSVKWKYNGEELKPSDRVKIENLDGQTQRLTVTNAKLDDMDVEYCE 4004
QY 1229 -----IYKQFVQKPCGRFPALNPVVGEMTQVRRRPERA-----1264
Db 4005 ASNEEGDVMSDVLTVKEPAQVAPG-----FPKELSAIQVKETETAKFECKVSGTKP 4056
QY 1265 -----IGPSQTH---TIRIQYSPARLAYQDKGVLH 1291
Db 4057 DVKWFKDGTPLEKDRVHFESTDDGTQRLVIEDSKTDQGNRYIEVS-----NDAGVAN 4110
```

```
QY 1292 NEVKVSI-----LWRGLPNV-----VTSA-----1310
Db 4111 SKYPLIVVSETLIKIKKGLTDVNTQGTKILLSVEVEGPKTKVWKYKGTETVTSQTTKI 4170
QY 1311 -----ISLPINRK-----1318
Db 4171 VQVTESEYKLEIESAEMSDTGAYRVVLSTDSFVSSESSATVTVTKAAEKISLPSFKGLAD 4230
QY 1319 -----PD-----1320
Db 4231 QSVPKGTPPLVLEVEIEGPKDVKWYKNGDEIKDKVEDLNGKYRLTIPDFOEKDVEYS 4290
QY 1321 -----GYDYAFSK-1329
Db 4291 VTAANEGATESAKVNVNSAKPELVISGLVPTTVKQGETATFNKVGKVPKVKYKNGKE 4350
QY 1330 -----DOYVNIIDVP-----SRTARAI-----1345
Db 4351 IPDAKTKDNGDGSYSLEIPNAQVEDAADYKVVNSNDAGDADSSAALTVKLADGDKVKP 4410
QY 1346 -----TTRSGQTLS---KV-----WY 1358
Db 4411 EIVSGLIPTTVKQGETATFNKVGKVPKVKQWKY 4443
```

RESULT 10
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.13) - yeast (*Saccharomyces cerevisiae*)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR0196
C;Species: *Saccharomyces cerevisiae*
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S48478; A26877; B26877; S27281; J06123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; G:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600; PMID:3106330
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:gl72522; PIDN:AAA35014.1; PID:gl72525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YAZ>
A;Cross-references: EMBL:M16165; NID:gl72523; PIDN:AAA35015.1; PID:gl72526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from *Saccharomyces cerevisiae*.
A;Reference number: S27281; MUID:89031230; PMID:3141213
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal growth of *Saccharomyces cerevisiae*.
A;Reference number: J06123; MUID:96323237; PMID:8710886
A;Accession: J06123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:gl304386; PIDN:AAC49609.1; PID:gl304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001456
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F;5-21/Domain: transmembrane #status predicted <TM1>
 F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 13.3%; Score 971; DB 1; Length 1367;
 Best Local Similarity 22.3%; Pred. No. 7.9e-13;
 Matches 377; Conservative 128; Mismatches 48; Indels 700; Gaps 76;

QY 13 LLSVFVIQVSSQDLSSCAGCGGYSRDATC-----NCDYN----- 49
 Db 6 LLAYLVLSLLFNSALGPTTALVPRGSGEGTSCNSIVNGCPNLDFNWHMDQOINQMTLDV 65
 QY 50 ----- 49
 Db 66 TSVSWQDNTYQITIHVKGENIDLKYLWSLKIIGVTPKGTQVQLGYNYNTYLIDPTD 125
 QY 50 -----CQHYMECCPDFKRVCTAELSCKGRCE----- 76
 Db 126 FTATFEVYATQDVNSCQVWM--ENFQIQ-----PEYLGSAQAQVASSWQWGT 171
 QY 77 SFERGREG-----DCDAQC----- 90
 Db 172 SFDLSTGNNYDNOGHSQTFPGYWNIDCNCNGTKSGTSTTSSESS 231
 QY 91 -----KKYDKCCPDYBSFCAEVHNPTSPSPSKKAPPPSGASOTIK 130
 Db 232 TTTSSTSESSTTSTSESSSTSTAP-----ATPTTSTCKEKP-----PTTT 278
 QY 131 STTKRSKPPN-----KKTKKVISEBEITEVDKNNKRTKKPTPKPPVDEAGSL 183
 Db 279 SCTKEKPTPHHDTPCTKKKT-----TTSKTKTKTKTTTPV----- 315
 QY 184 DNGDFKVTTPDTSSTQHNKYSTSPKLTAKPINPRPSLPNSDTSKTSITVNKEIT-- 240
 Db 316 -----TSSSITE-----SSAPVPT-----PSSSTSESSAPVTSSTSESS 352
 QY 241 ---VETKEITTTNKQTS--TDGKEKTTSAKETOSIEKTSADKAPTSKVLAKEPTKAEIT 295
 Db 353 SAPVPTSESSSTSESSAPVTSSTSESSAPVTSSTSESS--APV-----PTPSSST 403
 QY 296 TKGPA-----LITPKPTPTPKPEASTTPKEPTTT-----IKSAP-----TTPKEPAPT 342
 Db 404 ESSAPVTSSTSESSAPV-----SSTSESSAPVTSSTSESSAPVTSSTSESSAPV 459
 QY 343 TKSAPTTPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKAP 402
 Db 460 TPSSSTSESSAPVTS--TTSESSAPVPTSSSTSESSAPV-----SSTSESSAP 511
 QY 403 -----TTPKEPAPTTPKEPTP-----TTPKEPAPTTPKEPAPTTPKEPAPTTP 452
 Db 512 VPTPSSSTSESSAPA---PTPSSSTSESSAPV-----SSTSESSAPV-----PTPSS 560
 QY 453 KEPAPT-----KEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKAP 503
 Db 561 TTESSSTPVTSSSTSESSAPVTPS--SSTSESSAPVPTPSSSTSESSAPVTPSS 617
 QY 504 TPKKPSPTTPKEPAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKAP 562
 Db 618 TTESSAPVTS--TTSESSAPVPTPSSSTSESSAPVPT-----PSSSTSESSAPV 669
 QY 563 KEPAPTTPKEAPTTPKKLPTTPEKLAAPTTPKAPAPTTPPELAAPTTPPEPTP---TTP 619
 Db 670 TPSSSTSESSAPV-----SSTSESSAPV-----PTSSSTTES 717
 QY 620 EPAPTTPKAAANPTKEPAP-----TTPKEPAPTTPKEPAPTTPKEAPTTPKGTAPT 672
 Db 718 SSAPVPTPSSSTSESSAPVPTPSSSTSESSAPV-----SSTSESSAPVPTPSSSTTE 773
 QY 673 TLKEPAPTTPKPAKELAPTTTKEPTSTSDK---PAPTTPKGTAPTTPKEAPT----- 725
 Db 774 SSSAPVPTPSSSTSESSAPVPT--PSSSTSESSAPVPT--TPSSSNTTSSAPSSPTPSS 830
 QY 726 -----TPKEPAPTTPKGTAP-----TTLKEPAPTTPKPAKELAPTTTKGP-----T 768

Db 831 STESSVVPVPPSSSTTESSAPVSSSTTESSAPV-----PTPSSSNTSSAPSSIPFS 886
 QY 769 STSDKPAPTTPKETAPTTPKEPAPTTPKPAPTTPET--PPPTTSEVSTPTTKEPTTI 826
 Db 887 STTESFSTGIT--VTSSSKYPGSQTETSSTETIVTKTTSVTTFSTTTITTV 943
 QY 827 HKSDESTPELSAAPTTPKALENSKPEGVPPTKTPAATKPEMTT-----TAKDKTTER 879
 Db 944 CSTGNSAGETTSCSPKVTIT-----VPTTTTSTVTSTTTITTVTCSTGNSAGET 998
 QY 880 DLRTTPTET--TTAAAPKMTKETATTTTEKTESKITATTTQVTS--TTQDTTPFK-- 929
 Db 999 TSGCSPKIIITTVPCSTSPSETASESTTSTPTTPTVTVSTVTVTTEYSTSTKPGGEITT 1058
 QY 930 -----ITTLKTTTLA--PKVTT-----TKKTIITTEIMNKPEETAKEPKDRATNSKATTPK 977
 Db 1059 TFWTKNIPTTVLTITIAPTPSVTVTNFTPTITITVC-----STGNSAGET-- 1105
 QY 978 POKETKAPKPTSTKPKTMPRVKPKTTPTPRKMSTMP-----ELNPTSRIAEAM 1029
 Db 1106 -----TSGCSPKT-----VTTVPCTSTGTGEYTTAATLVTAV 1139
 QY 1030 LQT--TTRPNOTNSKLVEVNPKSEDAGAGEETPHMLLRPHVFMPEVTPDMDYLPRVP 1087
 Db 1140 TTVVTTSESTGNSA----- 1155
 QY 1088 QGIIINPMLSDETNICNGKEVDGLTLRLNGTLVAFRGHYFWMLSFPSPSPARRITEVWG 1147
 Db 1156 -----GKTTTGYTTKSVPT-----YVTTLAPSAPVTPATN-----A 1187
 QY 1148 IPSPIDIVTRCN-----CEGKTPFFKDSQYWRFTNDIKDAGYKPIFKGFGG 1195
 Db 1188 VPTIIT--TECSAANTNAAGETTIVCSAKT----- 1215
 QY 1196 LTGQIVAAALS--TAKYKNWP-----ESVYFFKRGSGSQIYKQEP----- 1234
 Db 1216 ----YVSSASAGENTAPSATTPVTALPTTVITTESVGVNSAGETTTGTTKSIPITYI 1271
 QY 1235 VQKCPGRPALNPVYXGEMT-----QVRRRRRFEAIGPSQTHTRIRIQYSPARLAYQ 1285
 Db 1272 TTLIPSGNAKNVETVATAPNISIKTSLQATNASASSVAP----- 1313
 QY 1286 DKGLVHNEVKVSIILWRGLPNVVTSAISLPMIRKPDGYDYAFKQDYNNIDVPSRTARAI 1345
 Db 1314 -----VWTS-----PSLTGP-----LQASGSVAV 1332
 QY 1346 TTRSGOTLSKVM 1357
 Db 1333 ATYSVFSISSTY 1344

RESULT 11

T42629

tenascin-X - bovine

N;Alternate names: flexilin

C;Species: Bos primigenius taurus (cattle)

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004

C;Accession: T42629

R;Elefteriou, F.; Exposito, J.Y.; Garrone, R.; Lethias, C.

J. Biol. Chem. 272, 22866-22874, 1997

A;Title: Characterization of the bovine tenascin-X.

A;Reference number: 22180; MUID:97426436; PMID:9278449

A;Accession: T42629

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-4135 <ELE>

A;Cross-references: UNIPROT:O18977; EMBL:Y11915; NID:G2462978; PIDN:CAA72671.1; PID:G246

C;Gene: TN-X

C;Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin typ

C;Keywords: extracellular matrix; glycoprotein; heptad repeat

| | | |
|---|------|--|
| Query Match | | 12.7%; Score 931.3; DB 2; Length 4135; |
| Best local Similarity | | 12.3%; Pred. No. 6e-11; |
| Matches 462; Conservative 197; Mismatches 496; Indels 2597; Gaps 125; | | |
| QY | 23 | SSQDLSS--CAGRC-GEYGRDATCNCDY-----NCQHYMEC-----56 |
| DB | 518 | AGEDGSRRCFGDCRGRCGDGVCSDVGEYEGDCGRKSPRCQCGQCGLEGRVCDD 577 |
| QY | 57 | -----CPD-----59 |
| DB | 578 | GYEGEDCGVRRCDRCNQRCQGVQGVCTCWEGFAGEDCGLRVCPSCNCHRRRCENGRCVC 637 |
| QY | 60 | -----FKVCTAELSCGR-----73 |
| DB | 638 | DSGYTGSCATRTCPAD--CRGRGRCVQGVCHVGYGEGDCGOEBPPASACPGCGCPRE 695 |
| QY | 74 | -----CPESFERGREG--DCDAOCKYDKC-----CPD-----99 |
| DB | 696 | LCSAGOCVCVEGF-RGPDCAIQTCPCDCRGREGCGSCVQCDGVAGEDCGEEVPAIEGM 754 |
| QY | 100 | -----YE-----SFCABVINPTS-----112 |
| DB | 755 | RMHLEBETTVRTEWTRAPGNVDAYEIQFIPTTEGASPPFTARVPSSASAYDQRLAPGQE 814 |
| QY | 113 | -----PPSSK-----117 |
| DB | 815 | QYTVRALRGTNWGPASKTITIMDGPQDLRVVAVPTTLELNLWLRQAEVDFVSYV 874 |
| QY | 118 | -----117 |
| DB | 875 | SAGNQRVRLEVPSEADGTLTGLMPGVEYVVTVAERGRAVSYYPASIRANTGSSLSGLG 934 |
| QY | 118 | --KAPPSGASQTKSTTKSPKPNKKTKKVIIESEIIEVKONKNRHK-----167 |
| DB | 935 | ATDBPPPSG--PSTTQGAQAPVLOQR-----POELAEVLGDKDKTGLRVAWTAQ 983 |
| QY | 168 | -----KPTPKPP-----174 |
| DB | 984 | PDFTFTHQLRLRVPEGGAHELLPGDVQALVSPPPESPYELSLRGIPSGPSAPLI 1043 |
| QY | 175 | ---VVDEAGS-----GLDNGDF-----188 |
| DB | 1044 | YQIMDKDGEKPGKPLAPPLRGKLTVDVTSLSLLHWTVEGEFDSFVIOYKDRDRPQV 1103 |
| QY | 189 | -KVTPDTSITQHN-----KVSISPKITTAQPI--NPRSLPN-----224 |
| DB | 1104 | VPVEGPQRSALISNLDVGRKYKFLVYGLVGRKRGHGLVAEAKILSQTDPSVTPPRLGNL 1163 |
| QY | 225 | -----224 |
| DB | 1164 | WVTDPTDLSLHSLWTVPEGQFDSFMVQYRDRAGRQVVPVREGPDRESVLIISPLDPHKYRP 1223 |
| QY | 225 | ---SPTSKESTSLVNKETTETKE-----TTTTNKQTS-----254 |
| DB | 1224 | TLFGIANKRHGPLTADGTTAPEKKEPRHPEPPERPLLGLTAVAGATADSLRLSWTVAQ 1283 |
| QY | 255 | -----TDGKEKTSAKETQSIEKTSADKLAPTCK-----VLA 286 |
| DB | 1284 | GSFDSFVVQYKDAQGRPQAVPTGDENEVAIPSLPEPRKYKMNLYGLHGRORVGPVSVVA 1343 |
| QY | 287 | KPTPKAETTTKGPALTTPKE-PTPTTPKEPASTTPKE-----TPTTIKSAPTP 335 |
| DB | 1344 | TTAQ-Q-EVLDETPSATMEETPSTPEPTKAPESPEKELLGELMVMVTGSSPDSLSLWTP 1402 |
| QY | 336 | K-----EP-----338 |
| DB | 1403 | QGHFDSFTVQYRDGQPRVNRVPGDEGVTISGLEPDHKYKMNLYGFHQRVQPMMSVI 1462 |
| QY | 339 | APTITKSAPTTPKE---PAPTTTKE-PAPITPKAPAPTTPKE---APTITKSAP---386 |
| DB | 1463 | GVTABEETPSTMEETPSTPEVEETESPMEPSTEAPAEPLLGELTWTGSSPDSLS 1522 |
| QY | 387 | ---TTPK-----EP-----392 |

| | | |
|----|------|--|
| DB | 1523 | LSWTVFQGHFDSFTVQYKGRDGPQVVRVGRGEETEVTIGLEPGRKYKMNLYGLHGQRSG 1582 |
| QY | 393 | -----APTTPKAPAPTTPKEPAPTTPKE-----TPTTPKE-----423 |
| DB | 1583 | PVSTVGVTAPQABETTPATEP-----PKEPRIGELTVDVTDVSTNSVGLWTVSEGOFDS 1635 |
| QY | 424 | -----PAPTTKECAPTTTPKE-----APTAPK 445 |
| DB | 1636 | FMVQYKDRDQCSHVVPVAADQREATVSGLEPERKYMNVYGLHGGQVGLSVVALLAPV 1695 |
| QY | 446 | KPAPT--PKEP-----APTTPKE-----PAPT 467 |
| DB | 1696 | PPDPVTPPVEPRIGELTVDVTDVPSVGLSWTVAEGEEDSLVQYKDRDGPQVVPVATD 1755 |
| QY | 468 | TKE-----PSPTT-----KEPA 480 |
| DB | 1756 | QREVTIGLEPFSRKYKFLFGIQDKRSTVSVEAKTVGRGDASPGAPRLGELMWTDPT 1815 |
| QY | 481 | PTTITKSAPT--490 |
| DB | 1816 | PDSLRLSWTVPEGHFDSFVQVKDRDGPVVSVEGHERSVTISPLDSGRKYRFLVYGLG 1875 |
| QY | 491 | TKEPAPTITKSAPTTPKEPSTTPKEPAP-----TTPK-----523 |
| DB | 1876 | KRRHGPLTTGTTTTRAVDEAGTKRSPKRLGEBELQVGTNSVGLSWTVPEGHFDSF 1935 |
| QY | 524 | -----BPAP 527 |
| DB | 1936 | VIQYRDGQGVVPVVEGSRREVSVGLDPARRYKLLYGLSRDKRVGPISAIATVPAP 1995 |
| QY | 528 | TTPKKPAPTTPKEPA-----PTTPKEPAP-----TTTK-----556 |
| DB | 1996 | REETKAEPAIFSPASPPLLEVTLEEAAPHLSLWTSATEGEFDSFEVQYTDENGLOQE 2055 |
| QY | 557 | -----PAPTAPK-----EPAPT-----568 |
| DB | 2056 | VNVGQDQHDITISLSDHRYLVSLYGFHDQVRGPAHIEAMTAPREDDPESLSSTQ 2115 |
| QY | 569 | TPKETAP-----TTPKLTTP-----TTP-----587 |
| DB | 2116 | TPSTAVPEPHIKPLGELAVTDTTTPDSLSLSWTVPQGFHFLIOYKNGDQPKVVRVP 2175 |
| QY | 588 | -----KLAP-----TTPKPAFTTPEELAPTTP-----P 613 |
| DB | 2176 | DESVTISGLEPDHKYKMNLYGFHNRQRMGVSVIGVTABEETP-----SPTEMEETP 2229 |
| QY | 614 | TPTTPEE-PAPTTPKAAAPNTPKP-----637 |
| DB | 2230 | SPTEMEETPSTPEVTEVPEPPEPFLBELTGTGSSPDSLSLWTVPGHFDSEFTVQYKN 2289 |
| QY | 638 | -----APTTPKE--PAPT 648 |
| DB | 2290 | GDGQPKVVRVPGHDKGVTISGLEPDHKYKMNLYGFHNRQVRGVPVSVIGVTABEETPST 2349 |
| QY | 649 | TPKE-PAPTTPKETAPTTPK-----GTAPTTL-----674 |
| DB | 2350 | EVEETPSTPESTEAPPEPPEPVLGELMVTGSSPDSLSLWTVPGHFDSEFTVQYKGRDG 2409 |
| QY | 675 | -----KEPAPTTPKPP 685 |
| DB | 2410 | POVVRVGEETEVTVEGLEPGHKYKMNLYGLHGRVRGVPVSTVAMTAREEPPASPPLK 2469 |
| QY | 686 | APKELAPT-----TTKE-----697 |
| DB | 2470 | QLGELTVDATPDSLSLSWTVPQGFHFLVQYKNGDQPKAVRVPGDEGVTISGLEPD 2529 |
| QY | 698 | -----PTS-----TTSK-----PAPTTPKGTAPTTPKE-----722 |
| DB | 2530 | HKYKMNLYGFHNRQVRGVPVSVGLTVSEKQDQEMTAPDTLPTAAPEPEIKPLGELVVD 2589 |
| QY | 723 | -----APTTPKEPAPTTPKGTAP-----740 |

Db 2590 ATPDLSLSWTVPEGQDFHFLIQKNGDQPKAVRVPQDGEDEVITISGLEPDHKYKMNLYG 2649
QY 741 -----TTLKE--PAPTPPK--PAKELAPTTTKGPTSTTSKPAPTTPKE 782
Db 2650 FHDQRMGPSVIGVTTABEETPGTEMEQTPSPTEVEET--PGPTE-VEETPSPTEPST 2706
QY 783 TAPTPKE----- 790
Db 2707 EAPPPPEPLLGELTVTGSSPDSLSWTVPGHFDSTIOHKRGDGPQVVRVGEETE 2766
QY 791 -----PAPTPPKAPPTTPTP- 807
Db 2767 TIGGLEPGRKYKMNLYGLHSGQVGVSTGVTDQEVVEETPSPTEPSTEAPPPPEPL 2826
QY 808 ----- 807
Db 2827 LGDLTVTGSSPDSLSWTVPGHFDSTIOYKGRDRPQVVRVGEETEVIIGGLEPRHK 2886
QY 808 -----PPTTSEVSTP---TTTKBPTTIHKSPDESTPELSAEP----- 841
Db 2887 YKMNLYGLHGGRRVGPVSTGVGTAPDYDAMTTQTPST-----SVPEPTPKPRLGEVTV 2939
QY 842 --TPKAL-----ENSPK-----EP----- 853
Db 2940 TDA TPDSLSWTVPEGQDFHFLVQKNGDQPKAVRVPQDGEDEVITISGLEPDHKYKMNLY 2999
QY 854 -----GVPTT--KTPAAKPEMTTAKDKTTERDLRTTPTTTAAPKMTK 896
Db 3000 YGHDRGVGPVIGVTTABEETPSPTEP-----ETPSPTEME 3039
QY 897 ETATTTEKTES-----KITAT----- 913
Db 3040 ETPSPTEPSTEAPPPPEPLLGELTVTGSSPDSLSWTVPGHFDSTVQYKGRDGPQV 3099
QY 914 -----TTQVT----- 918
Db 3100 VRVGGEETEVTVGLEPGRKYKMNLYGLHGGRRVGPASTVGTASLTATERPLAPRLGELA 3159
QY 919 -STTTQDTT-----PF-----KITLTKTTILAP----- 940
Db 3160 VAVVTSDBLSWTVGEGPDSFLVQYKVGQDQVAVPDVAGDUREVTVSSLPAGRKYKFL 3219
QY 941 -----KVTTTKTITTTIMNKPEETAKP----- 964
Db 3220 LFLGRDEKRHPVSADAKTLPTD-----KPAERLGLTVDTPGSGVGLSWTVPE 3269
QY 965 -----KDR-----ATN 970
Db 3270 GEFDSFMVQYKDRDQGPVAVADQREVTVPGLEPNRKYKFLYGLVGRKRLGPISAEG 3329
QY 971 SKATTPKPKP----- 981
Db 3330 STAPLEKERPPRLGELTVTDETPNSRLSWTVAGQRFDSFVVQYRGDGPQRMVAVAA 3389
QY 982 -----TKAPKKP----- 988
Db 3390 DOREFTVEGLEPGRKYKFLYGLLGGQRLGPASVLGWTAPEDTPAPWHAATEAPKPEG 3449
QY 989 ----- 988
Db 3450 FRLGVLAVRDVS PDSLSWSVVGPPDSFVVOYQDTPDQOPALLVGGDQNKVLVSGLEP 3509
QY 989 ----- 988
Db 3510 STSYEFFLYGLHGGKRLGPVSAEGTTGPVAGOTPGEPGRLSHLSVTDVTTSSLRNWE 3569
QY 989 -----TSTKKPKTMRVRKPKTTPTPRKWTSTMPELNP----- 1021
Db 3570 APPEAFDPLRFGVPSFSTLEQLRELTVPGTRR--SAVLRDLHFGTLYTLTYLG 3628
QY 1022 -----TSRIAEAML-----OTTRPNOTNSKLVE---VNPXSEDA 1054
Db 3629 LRGPBKADSIQGTARTLSPVLESFRLQFSEIRTSARVSWTPTTSRVDGFKVSYQLADG 3688

QY 1055 G-----CAGCETHMLLRPHVFMPEVTPDMDYLRPRVNOQIIINPM 1095
Db 3689 GEPOSQVQDGRTOKLEGLIFGAQIEVTVSVRG---FEESEPLTGFITVDP-----GPT 3740
QY 1096 -----LSDETNICNGK-----PVDGLTTLN- 1116
Db 3741 HLRALNUTDSALUHWKPPQTPVDYDVKVTA PGAPSLQASAPGSAVDYPLQGLVTHNTY 3800
QY 1117 -GTLVAFRGRHYF-----WMLSPFSP- 1137
Db 3801 TATLRGLRGNFTSPASITFTTGLEAQDLEAKVTPRTALLTWTAEVSTPGYLLSNT 3860
QY 1138 PARRITEV-----WG-----IPSPID 1153
Db 3861 PGGTQBIILLPGGVTSHLRGLFFSTPYSTWLRAMGDSFTPPVSTSTFTTGLRIPPRD 3920
QY 1154 -----TVFTRCNCE-----GKTFEFDQSYW 1174
Db 3921 CGEEMQNGVSTRTTITFLNGNRERPLNVFCMDTGGGWLVFQRRMDGKTDFWRD---W 3977
QY 1175 RFTNDIKDAGYKPKIFKFGSLTGQI VAALSTAKYKNW--PESVYFFRKGSIQOYIYKQ 1232
Db 3978 -----EDYAH-----GFGNISGEF-----WLGNEALHSLTKAGD----- 4006
QY 1233 EPVOKCPRRPAALNYPVYGEWTVRRRFRFRAIGPSQTHIRIQYSPARLAYQDKGVLHN 1292
Db 4007 -----YSLRVD-----LRAGEAVFAQ 4023
QY 1293 EVKVSILWRGLPNVVTSAISLPNIRKPDGYD-VYAFSKDQYVNDIVPSRTARA---ITTR 1348
Db 4024 -----YDSPQVDSADEYRULHLEGYHGTAGDSMSYH 4054
QY 1349 SGQTL-----KVWY-NC 1360
Db 4055 SGVSFSARDRPNLLISCASVYRGAWMYRNC 4086

RESULT 12

S49915

extensin-like protein - maize

C;Species: Zea mays (maize)

C;Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004

R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A;Description: Pex genes: pollen-specific genes with extensin-like domains.

A;Reference number: S49915

A;Accession: S49915

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1188 <RUB>

A;Cross-references: UNIPROT:Q41805; EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g6001

Query Match

12.5%; Score 917.8; DB 2; Length 1188;

Best Local Similarity 18.9%; Pred. No. 7.9e-12;

Matches 299; Conservative 81; Mismatches 397; Indels 802; Gaps 57;

QY 10 LLLL---LSVFVITQVSSQDLSSCA-----GRGCE----- 36

Db 17 LLLLAACLACSQAVTSAEASVIAHRQLLAMKEAGGEAGDLPADFEFDDRVGAANFPN 76

QY 37 -----GYSRDATCN-----CDYNOCHYMECCPDFKRVCTAEL 68

Db 77 PRLRRAYIALQAWHRAFYSDPKGY---TANWVGEDVCKYNGVICTEALDDPKITVVAGI 132

QY 69 SCKGR----- 73

Db 133 DLNGADIAGYLPPELGLLTLAPFFHINTNRFCGIIPKMSRSLSLLEHFDVSNRRFVGVPF 192

QY 74 --CFES-----FERGRECCDCA----- 88

Db 193 YVCLEWVSLKYLLDRNFDEGELPPALFDK---DLDAIFVNTNRVFGPIPENLGNSTAS 248

| | | | | | | | |
|-----------------------|------|--|-----------------------------------|-----------------------------------|----------------------|--------------|--------------|
| Query Match | | | | 12.5%; | Score 916.8; | DB 2; | Length 5762; |
| Best Local Similarity | | | | 11.3%; | Pred. No. 2.5e-10; | | |
| Matches 535; | | | | Conservative 197; | Mismatches 474; | Indels 3512; | Gaps 155; |
| Qy | 2 | AWKTLPIYLL--- | LLSVFVIQVSS----- | QDLSS 29 | | | |
| Db | 753 | AWHSVTDMFLSPVDLQIFRSTQSKYKTIKHEDLALITIPESLSDGSLTFQEDLLQ 812 | | | | | |
| Qy | 30 | CAGRCGGYSDATCNDYCNQHYMECCPDFKRVCTAELS----- | CKGRCF----- | ES 77 | | | |
| Db | 813 | PIDSTGQG----- | ----- | | | | |
| Qy | 78 | FEREGCCDAQCKKYDKCCPDY----- | ESFCAEVH-NPTSPSSKKAPPS 123 | | | | |
| Db | 852 | TYSNQVDLSSTHLKTSLELDPDYTWGLEPSLYQOITQSSPKSLHENPKSFP----- | 902 | | | | |
| Qy | 124 | GASOTIKSTTKRSKPKPNKKX----- | | 144 | | | |
| Db | 903 | ---VLKPSAAQALKPRKEKLSPTNNMIPHHPSKPLKNMVTHIPAHKMTVPRQIQEDQG 958 | | | | | |
| Qy | 145 | ----- | TKVIESEBEITEVK----- | DNKK 162 | | | |
| Db | 959 | EYTISSNGSQPLDLEVLTSGLIPEVKHILPKRTVNPQYSQVKISHSQHVETQHENSE 1018 | | | | | |
| Qy | 163 | NRTKK----- | KPTPKPPVVDEAGSLDNGDFKVTTPDTS----- | 197 | | | |
| Db | 1019 | TTIVQPLDLEFALNQLTPKE----- | NFAQTQDITTMIGPPEKVEIAQPE 1065 | | | | |
| Qy | 198 | ----- | ----- | TOHKK 202 | | | |
| Db | 1066 | HHEGTIPIQDOAEYSTLPTVSFQPLDQLBITTSEAIRBPHPTVPQQIIVHPPHPL 1125 | | | | | |
| Qy | 203 | VSTSPKITTAKP----- | INPRPS----- | 220 | | | |
| Db | 1126 | VIHSEQVYTOHPNTEALIQPLDLELITPQPTAEGELPQLDQSTTQIIEPTVVVGPV 1185 | | | | | |
| Qy | 221 | ----- | LPN----- | SDTSKETSITVNKETTET----- | KETTITN----- | 250 | |
| Db | 1186 | PIYEEVTVQITSQDAEYRPSPTVSFQSLDLELITISPEATREYHPSLLQOITIVNPPH 1245 | | | | | |
| Qy | 251 | ----- | KQSTSD----- | GKEKTTSAKETQSIKTSAMDLAAPTQKVL 286 | | | |
| Db | 1246 | PLVIHSEQVHTQHPNLTEATVQPLDPLTITPQPTTEGELPQLQDS----- | TSQIII 1298 | | | | |
| Qy | 287 | KPIPK----- | AEITTKGPALTTP----- | KEP-TPITPKPEAST 318 | | | |
| Db | 1299 | EPPIKVVALVPVYQVQSQDAEYTTSTVSFQPLDQELTITSAIRBPHPTVPQOITIV 1358 | | | | | |
| Qy | 319 | TP----- | KBPTPTTIKSAP----- | TTPKAPA----- | PTTKSAPTTPKEPAPT 356 | | |
| Db | 1359 | HPTKHPLVIIHSEQTOHPNTEVTVQPLDLELITWPTQPTAEGELPQLDQSTTQIIEP-PT 1417 | | | | | |
| Qy | 357 | TTKEPAP----- | TPPKE----- | PAPTT----- | | | |
| Db | 1418 | VVVGPPVPIYEEVTVQITTSQDAEYRPSPTVSFQSLDGLTITPEPTEHFITQKTIVFPBM 1477 | | | | | |
| Qy | 374 | ----- | TKEPAP----- | TTTKSAPTTPKEP----- | 392 | | |
| Db | 1478 | YTDVTLPOQVSVQHLKPTGIVQPLDLELITPQPTPEGELSQTVQESTQNKEPHKEVV 1537 | | | | | |
| Qy | 393 | APT----- | TPKK----- | | | | |
| Db | 1538 | APVVPVQAVTVPTPSQVQAEYQKSLQPLDLELITVTSPTKEAVHSTISKNLSAINPQYVH 1597 | | | | | |
| Qy | 400 | ----- | PAPTT----- | PK----- | | | |
| Db | 1598 | IOHPNPAEATVQPLDLELITSSSQPTAEGELLYSMQETVQISEPPKQVVPVPEYQEV 1657 | | | | | |
| Qy | 408 | ---PAP----- | TTPKETPTTPPKE----- | 423 | | | |
| Db | 1658 | AVPAPVQDAKYPLSSIVSLNSLDQELTSLSELLGEAHLQITTPDETMTWLPKPKDROGIYD 1717 | | | | | |

| | | | | | | | |
|----|------|---|-------------------------------|-------------------|--------|--|--|
| Qy | 424 | ----- | | | | | |
| Db | 1718 | HDHKKHLNLTEVTNQPFHLEHTVQHQPTIEBERSQSIQKKTQTQITEPGKKVPLAQESEE 1777 | | | | | |
| Qy | 424 | ---PAPTTKEPAPTTP----- | | | | | |
| Db | 1778 | VTIPMPILKETAPFTPHSMALQSLDEKLTIHSHSPGWTQOAHNKKESKGHTTGKILLDYA 1837 | | | | | |
| Qy | 437 | ----- | | | | | |
| Db | 1838 | EPANMEIELKHGLFLKLTTEATTESENTNQMTSLKQVLTALFTQNKKSMLPALVESQDES 1897 | | | | | |
| Qy | 438 | EPAPTAEPKA----- | PTTPK----- | EPAPT----- | | | |
| Db | 1898 | QPPNMSLQPLDQELTSSQPHGMVPHIPNTPEKILYHYAEP-PTGPFVPEPPELFLKTT 1956 | | | | | |
| Qy | 459 | ----- | TPKE----- | PAP----- | | | |
| Db | 1957 | KSRPVQGTATQMAASPKEVMYRAPENKEAVLSGEGEDQDESPPNMSLQSLDQELTSS 2016 | | | | | |
| Qy | 470 | EPSPTTPKEP----- | APT----- | TTKSAP----- | | | |
| Db | 2017 | QPHGWIPHPNTHGKIYLYHYAEPPTGPFVBPDLFLKTTKSKPEVWTLRTDKSRKEMV 2076 | | | | | |
| Qy | 489 | ----- | TTTKEP----- | | | | |
| Db | 2077 | SQSPKYEEAVLPVHGQEEESRPPNMSLQSLQELTSSQPHGMVPHHPNTHGKIYLYH 2136 | | | | | |
| Qy | 495 | ---APT----- | TTKSAP----- | | | | |
| Db | 2137 | AEPTGPFVBPDLFLRTTKSKPVQGTATRMVKSPEEMVSLDPENKEAVPPAQGEKGE 2196 | | | | | |
| Qy | 504 | ----- | | | | | |
| Db | 2197 | SPSSPNMSLQSLDHELFMSQPHGWIPHPKTPDKIYLYHYAEPPTGPFVBPDLFLRTT 2256 | | | | | |
| Qy | 506 | PKESPSTTKKEPAPTTPKSPAPTT----- | KKPAP----- | | | | |
| Db | 2257 | KSKPVQGTITTEMA-KSPKEMVSQTPYEKAVLSGEGEDQDESPPNTSLKSLDQEVAMS 2315 | | | | | |
| Qy | 536 | TTKEPAPTPK----- | BPAP----- | TTTK-KPAPTAP----- | KE 564 | | |
| Db | 2316 | SQHSQVPHPPKTPGKIYLYHSIEPPGPFVKPTDILLVKTTKSKPAEWTPRIDKLLKE 2375 | | | | | |
| Qy | 565 | PAPTTPK-ETA----- | | | | | |
| Db | 2376 | MVPHSPYEVEAVPPAHGEGQDESPPNMPLQPLDQELTSSQPHGMVPHHPNTPGKIYL 2435 | | | | | |
| Qy | 582 | ---TPTTPEKLAP----- | TTPEKAPTTPEELAPTTPEEPTTP----- | | | | |
| Db | 2436 | HYAEPPTGPFVBPDLFLRTTKSKREVQSPKEIA-KSPKEMVSQTPYEKAVLSGEGED 2494 | | | | | |
| Qy | 619 | ---EAPAP----- | TTPKAAAPNTPK----- | BPAP----- | | | |
| Db | 2495 | QDESPPNMSLQSLDQEVMTSSQPHSGVPHPKTPGKIYLYHSIEPPGPFVKPTDILLV 2554 | | | | | |
| Qy | 640 | ---TTPKAPAPTT----- | KEPAPTTPK-ETA----- | | | | |
| Db | 2555 | KTTTKSKPAEWTPRIDKLLKEMVPHSPYEVEAVPPAHGEGQDESPPNMPLQPLDQEL 2614 | | | | | |
| Qy | 663 | ----- | PTTP----- | | | | |
| Db | 2615 | TLSSQPHGMVPHHPNTPGKIYLYHYAEPPTGPFVBPDLFLRTPKSKPVQGT-PTQMAKS 2673 | | | | | |
| Qy | 675 | ----- | KEPAP----- | | | | |
| Db | 2674 | PSEMVSLSFKNKETVFPAGCKQDESPPNLSLQSLQDEITMSSQPHGWIHPNTHGK 2733 | | | | | |
| Qy | 680 | ----- | TTPKP----- | APKEL----- | | | |
| Db | 2734 | IYLYHYAEPPTGPFVBPDLFLKTTKSKFMQGSPRQIDKSPKEMFTQSPYBESLLPAHA 2793 | | | | | |
| Qy | 691 | ----- | AP----- | | | | |

```
Db 2794 EGQESRAPHFSLQDLQELSLSSHGHGWIHPHPNTPDKIYLHYABPTGPFVEPPDLF 2853
QY 693 ---TTTKEP---TSTTSK-----PAP-----708
Db 2854 FLTKTSKPVQGTATKTDSPEDRVSTQTPYKEAVLSGPGDQDESPPNMSLSQLDQE 2913
QY 709 ---TTPKGTAPTPKEP-----APT---T 726
Db 2914 LAISSQPHGWIHPSPNAPDKIYLHYABPTGPFVEPPDLFLTKTKSKPLQGTQMAKS 2973
QY 727 PKEPATPK-----736
Db 2974 PKEVSTQPYKEADISAPGENDESPSPNMSLHPLDQELSLSSQPHGWIHPHPNTHGK 3033
QY 737 ---GTAP---TTLKEP-----746
Db 3034 IYLHYABPTGPFVEPPDLFLTKTKSKPVQGSQIDKSPKEVFSQSPSESVLPAQA 3093
QY 747 ---AP-----748
Db 3094 EGQESRAPHMSLQLDQDLTLSSHGHGWIHPHPNTPGKIYLHYABPTGPFVEPPDLF 3153
QY 749 ---TTPKPK---APKEL---APTTKG-----766
Db 3154 FLTKTSKPVQGSQPRQVDKSPKEMFTQSPYEESVLPQAAGQESRAPHMSLQPLDQD 3213
QY 767 ---PTS---TTSKPAPTP-----780
Db 3214 LTLSSHGHGWIHPHPNTPDKIYLHYABPTGPFVEPPDLFLTKTKSKPAQTPTQMAKS 3273
QY 781 ---KETA-----784
Db 3274 PEMVSLSPENKETVFPQAQGOESISPPHMSLQPLDQDLTPSSHGHGWIHPHPNTHGK 3333
QY 785 ---PT---TPKE---PA---792
Db 3334 IYLHYABPTGPFVEPPDLFLTKTKSKPVQSVPRQIDKSPKEVFTQSPYEESVLPQA 3393
QY 793 ---792
Db 3394 EGQESRAPHMSLQPLDQDLTLSSHGHGWIHPHPNTHGKIYLHYABPTGPFVEPPDLF 3453
QY 793 ---PTTPKKA-----800
Db 3454 FLTKTSKPVHGSQPRQIDKSHKEMFTQSPYEESVLPQAAGQESRAPHMSLQPLDQD 3513
QY 801 ---PTTPEP---PPTSEVSTP---TTTK---EPTTIHKS 829
Db 3514 LTLSSHGHGWIHPHPNTPDKIYLHYABPTGPFVEPPDLFLTKTKSKPVQGSQPRQIDKS 3573
QY 830 PDE-----STPLSAEP-----841
Db 3574 PKEMFTQSPYEESVLPQAAGQESRAPHMSLQPLDQDLTLSSHGHGWIHPHPNTPDK 3633
QY 842 ---TPKALENSPKE-----852
Db 3634 IYLHYABPTGPFVEPPDLFLTKTKSKPVHGSQPRQIDKSPKEVFTQSPYEESVLPQA 3693
QY 853 ---PGVP-----856
Db 3694 EDQESRAPHMSLQPLDQDLTLSSHGHGWIHPHPNTPDKIYLHYABPTGPFVEPPDLF 3753
QY 857 ---TTTKT-----PA-----862
Db 3754 FLTKTSKPVHGSQPRQIDKSPKEVFTQSPYEESVLPQAAGQESRAPHMSLQPLDQD 3813
QY 863 ---ATKP-EMTTTAKDKT-----876
Db 3814 LTLSSHGHGWIHPHPNTHGKIYLHYABPTGPFVEPPDLFLTKTKSKPVQWTPTQIDKS 3873
QY 877 TERDLRTPE-----886
```

```
Db 3874 REEMVQSPENEEADIPRHGDQDELNRNPFIIQLDQELTLSSQPHGWVPHPPSTPDK 3933
QY 887 ---TTTAAP-----KMTKETATTE-----903
Db 3934 IYLHYABPTGPFVEPPDLFLTKTKSKPVQGTPTOLAKSPKEMVFTQTPYKEAVLSAPG 3993
QY 904 -----903
Db 3994 EDOETPSSPNMSLQDLQELTMSSQPHGWLPHPPNTHGKIYLHYADPTGPFVEPPDLF 4053
QY 904 ---KTTSK-ITATTTQVSTTTQ-----923
Db 4054 FLTKTSKPVQGTVTQMDKSPKQGLLPAHGEAQDESPPNMSLQPVQDQLSLSSQPHGW 4113
QY 924 ---DTPPKITT-----LKT-----936
Db 4114 ITHPPNTPGKIYLHYABPTGPFVEPPDLFLTKTKSKPLAGTPTQMAKSPKEMFSOTLE 4173
QY 937 ---TLAP-----940
Db 4174 HKEAVLSAPGEDQDESPPSSNMSLSQLDQDITLSPOQYGRIPHPVTSQKIYLHYVEHPT 4233
QY 941 ---KVTTKK-TITTEIMNKPE-----960
Db 4234 VPLFQHPNMFSTKTTTSKPLQLTQTOMAKSPKEIVSLRPEYKEAILPAQVESQESVLP 4293
QY 961 ---TAKPKD-----966
Db 4294 NMSLSQLDQELTLSSQPHGWIHPHPNTPDKIYLHYABPTGPFVEPPDLFLTKTKSKPV 4353
QY 967 -----966
Db 4354 QWTPTQIDKSLKXVAGSPYEBAVPAHGEQDESRYAENISLSQLDHDITLSSQPHGW 4413
QY 967 ---RATNSKAT--TPKP-----978
Db 4414 IPHPNTPDKIYLHYABPTGPFVEPPDLFLTKTKSKPAQWTPTPMAKSPPEMVSLSK 4473
QY 979 ---OKPT 982
Db 4474 NKEAVLPAQKDEESISPPNMSLQPLDQELTLSSQPHGWVHPNTHGKIYLHYABPT 4533
QY 983 ---KAPKK-----987
Db 4534 GPFVEPPDLFLTKTKSKPVQGTITQMVKSPKQGLLPAHGEAQDESPPNMSLQPVQDE 4593
QY 988 ---PTST---KPKXTMPVRKP---KT---TTT-----1008
Db 4594 LSLSSQPHGWIHPNTPDKIYLHYABPTGPFVEPPDLFLTKTKSKPLLGTTQMAKS 4653
QY 1009 PRKMTSTMPBLNPTSRAEAMLOTTTRPNOT-----PNSKL-----1044
Db 4654 PKEMVSTQPYKEAVLSGEDLDESTSPNMSLQPLDQELSLSSQPHGRIHPVTSQGIY 4713
QY 1045 ---VEYNPKS-----EDAGGA 1057
Db 4714 LHYABPTGPFVEPPDLFLTKTKKAPVQGTATHVEKSPKEMVFRSPYKQAVLPAHGEA 4773
QY 1058 EGET---PHMLL-----RPHVEMPE-----VTPDMYL 1082
Db 4774 QDESRSPPNMSLQDLQELTMSSQPHGWIHPHPNTHGKIYLHYABPTGPFVEPPDLFL 4833
QY 1083 PR-----VPMOQ-----IINPM-----1095
Db 4834 KTKSKPVQGTATQMAKSPPEMVSLLENKEAVLPAQGDGLEESIFPPNMSLQDLDELI 4893
QY 1096 LSDE-----TNI-----CNKPKVQDGLTUTRNGTLV 1120
Db 4894 LSSQPHGWIHPNTPSNILYHYABPTGPFVEPPDLFLTKTKSKPVQGT-----SPTLI 4948
QY 1121 A-----FRCH-----YFWMLSPEP 1135
Db 4949 AKSPREMSRSPYKQALLPGHGEKNSRAPNMSLQPLDQELTMSSQPYGW-----5001
```

QY 1136 FSPARRITEVMGIPSIDIVTRCNCEGK-----TFFFKDSQ-----Y 1173
Db : : : : :
QY 5002 -----IHPHNT-----PCKIYLHYAEPPTGPFVEPPDLFFLTKTKSKPVQ 5042
Db : : : : :
QY 1174 WRFTNDIKDAGYPRIFKGFGLTGQIVAAALSTAKYK----- 1210
Db : : : : :
QY 5043 W---TSAEIAKLPEIVS-----QTQKHESVLRAPEQHKNKSPSPNVSLQ 5086
QY 1211 -----NW-----PESVYFFKRGSGIOQYIYKQ 1232
Db : : : : :
QY 5087 PLDQELSLFSOPHGWIPHSNTPGKIYLHYAEPPTGPFVEPPDLFFLTK 5138
QY 1233 EPVQKCP-----GRPA----- 1244
Db : : : : :
QY 5139 KPVQGTPTMTKSPQEIIVSQSPGMEAGPTTKVEQDESPPSHASLOTLDOELTLSSQP 5198
QY 1245 -----LNY-----PVYGMTQVRRRRFERAIG 1266
Db : : : : :
QY 5199 HGWIPHPNPTAGKIYLHYAEPPTGPFVEPPDLFFLTKTKSKPVQGTATQSTQ----- 5250
QY 1267 PSQHTIRIQVSPARLAYQDKVLHNEVKVLSILWRGLPNVVTSAI-----SLPNIRKP 1319
Db : : : : :
QY 5251 -----SPKEI-----ISPSTRYKESVLTAADVQDESQSPPNISLQ 5286
QY 1320 DGYDYAFSKDQYINIDVPSTRARAITRRSGOTLSKVM 1357
Db : : : : :
QY 5287 PLDQELTLSSQPHVLISHPSST-----NANTGHTLGLKIY 5320

RESULT 14
T25697
hypoetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <FUL>
A:Cross-references: UNIPROT:Q94185; EMBL:U67956; PIDN:AA807691.1; GSPDB:GN00028; CESP:FI
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP.F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 12.1%; Score 882.5; DB 2; Length 1229;
Best Local Similarity 23.7%; Pred.No. 4.8e-11;
Matches 375; Conservative 107; Mismatches 377; Indels 725; Gaps 79;

QY 2 AKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSRATCNCYDNCQHYMECCPDFK 61
Db : : : : :
QY 4 AW-VVSFAFLILGNVQSSLLSKTINSQSR-----DFK 37
Db : : : : :
QY 62 RV-----CTAELSCGRCFSEFREGRCDDCAQCKYDKCOPVYESFCAEVHNTSP- 113
Db : : : : :
QY 38 IVKHKKNTCT-----CSC-----KCPD-----APSNFF 62
Db : : : : :
QY 114 -----PSKXKAPPSPGAS-----CTIKSTYKSPKPNKKT 145
Db : : : : :
QY 63 DVSTTISSINNDVIGPSGDS--NPTGSSWFQEIATVGGQTVKS-----E 107
QY 146 KKVISEEITEVKONKNRTRKKTPKPPVDEAGSLDNGDFKVTPTDITSTQ---HNK 202
Db : : : : :
QY 108 HNIIDSSVEV-----EKKVT-----TSTDASTINAPTQK 136
QY 203 VSTSPKI-----TTAKPINRPSLPSPNSDTSKETSATVNKET 239
Db : : : : :
QY 1058 RA-----OPTN-----EMDKEMEFKIR 1076

Db 137 DSTTPEIITGIIVINSKSESVDMSSTRFSTTSLP-----TTPELLTSPT 181
QY 240 TVEKETTITNKQSTDGKEKITSAKETQSIKTSKDLAPTCKVLAKTPKAETTKGP 299
Db : : : : :
QY 182 LVSTDSSTST-EQTSPONTTEIASPMET-----NTTE-- 213
QY 300 ALTPKBPPTT---PKEPASTTKEPTPTTIKAGPTTKEP---APTTKSAPTTPKEP 353
Db : : : : :
QY 214 ATTTVSFVSTLASEDETTVTAESTTTTIAEVSTTTTETPTTAESTTKSTT---K 269
QY 354 APTTKPAPPTTKEPAPTTTKEPAPTTTTSKAPTTPKEPAPTT---KKPAPTTPKEP 409
Db : : : : :
QY 270 APATTEPTPTTEE---VTTEASTTTSSETSEK---ETTLIDNKIAGPATG--- 321
QY 410 PTTTKEP---TPPTTKEPAPTTTKEPAPTTT---KEPAPTAPKKPAPTTTKEPAPTTTKEP 464
Db : : : : :
QY 322 PETTHFPVTGTP-----NFDATETPFVAKSDEKMTLSKTAATETTQOTTEV-DOPE 374
QY 465 PTTTKE-----PSPTTKEPAPTTTTSKAPTTP---TTTKEPAPT---TTKSAPTTPKE 508
Db : : : : :
QY 375 KEITKNVSIPIPTTVLVELVETTTSTTAKESDGFHTTLKLVTTADSDSTESATTVKPF 434
QY 509 PSPTTKE---PAPT-----TPK-----EPAPTTPKKPAP----- 535
Db : : : : :
QY 435 NEETTKSHVVPRKTKGTGVKVPKLELSFDEPTEIT-KAPHGKGLLEKTYHFVLSDNF 493
QY 536 -----TTTKEPAP-----PTTKEPAPT-----PTTKEPAPT 552
Db : : : : :
QY 494 ARYSEAKENDDYNHLDYNHYREAKEPTTTEESSTTEBVTTEBPANTGNPTTEN---PT 550
QY 553 TTKPAPTAPKEA---PTTKEAPTTPKKLTPTTPEKLAPTTPKEPAPTTPPELAPTTP 610
Db : : : : :
QY 551 TTEQPTSTABSTTALPFTTTEQTV-----TTEE---PTTAEKSTATQ---KPTTT 594
QY 611 EEPPTTPEPAPTTPKAAAPNTPKBPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP---TTPK 667
Db : : : : :
QY 595 QESVST---EKTSTTKA---STEE---PTTDEPTTTT---ESSTTGKATPELSSTSE 643
QY 668 GTAPTTLKEPAPTTPKKPAPKELAPTTPKEPTTSDKAPAP---TTPKGTAPTTPKEP 723
Db : : : : :
QY 644 ETTTTELK---ITTE-----GSTTTEPTTIAFAEASTGLIITDETTTSTTSTPE 692
QY 724 PTTKEPAPTTPKGTAPTTLKEPAPTTP-----KKPAPKELA 760
Db : : : : :
QY 693 ITSTKE---IVTESAITQTSVSVVSVESSTPQOLPERKAIIVNFKHNLVLKKEKRLLEKE 750
QY 761 PTTTKEP-----TSTTSKAPAPTTPKETAPTTPKEPAPTTPKAPAPTTP 804
Db : : : : :
QY 751 STSTGSDSSETTVVAENIDEVTTTEKVKVQVTTTTEKSTTQEEPTTTTTTTEKTTTS 810
QY 805 E---TPPPTTSE-VSTPTTKEPTTIHKSPDESTPELSAEPPTKALENSPKPGVPTTKP 861
Db : : : : :
QY 811 KTTTEKPTTSESATTTETTSEPT---TEST-----TVDTS 843
QY 862 AATKPEMTTAKDKTTERDLRTPETTTA----- 890
Db : : : : :
QY 844 SATTEESATTAETTTISAE---TSETTSESAAFITGESPENTALQSSQKGEENESSAE 900
QY 891 -----APKMTKETATTTETTESKTIATTTQ-VTST---TQDTPPKLTTLKTTTL 938
Db : : : : :
QY 901 KPGARDFVPKHKTTVPKPAETTSAAVASTTTEPTTEKSTTLETTTLETTTLETTTLETT 960
QY 939 APKVT---TTKKTITTTTMMNKPEETAKPKORATNSKATTPKQ----- 979
Db : : : : :
QY 961 PAFVTGAPVDETTINTLELLSK-----INNTOISQKPTDISKTALSSLSISGLI 1010
QY 980 -KPTKAPKKPTSTKPKTMDPRVKPKTTPTRKMTSTMPELN-----PTSRIA 1026
Db : : : : :
QY 1011 GSFTKAPMPTI-----HTTTDAAFVATATEASLNDGSDKKIIDEAQPTDEIR 1057
QY 1027 EAMLQTTTTRPQTPNSKLVEYNPKSEDAAGAEFTPHMLLRPHVFMPEVTPDMQYLPRVP 1086
Db : : : : :
QY 1058 RA-----OPTN-----EMDKEMEFKIR 1076

| | | | |
|----|------|--|------|
| QY | 181 | -----SGLDNGD----- | 187 |
| Db | 406 | EQPQLTSTCHIALNSNENLLGKQFQCTDSGEPLLPSTSEFGNVFFSAQNAAKQPSDKC | 465 |
| QY | 188 | -----FKVT-----TPDTSTTOHNVST-----SPALITAKPI | 215 |
| Db | 466 | SASPLLRQCIRENGNVAKTPRNTYKMTSLETETETPTSKTVSVARSGRSTFRNI | 525 |
| QY | 216 | NRP----- | 219 |
| Db | 526 | QKLPVESKSEETNTEIVECILKRGQKATLLQORREGEMKEIERPPFYKENIELKENDEK | 585 |
| QY | 220 | -----SLP-----PNSDTSKETS | 232 |
| Db | 586 | MRAMKESRTWQKCAPMSDLTDLKSLPDTLMKMDTARGQNLQTQDHAKAPKSEKGIKTX | 645 |
| QY | 233 | LTVVK-----ETTVETK-----TTTTNKQTSIDGKE | 259 |
| Db | 646 | MPCQSLQPEPINTPHTTKQQLKASLGKGVKEELLAVGKFTRTSGETTHTRBPAGDGKS | 705 |
| QY | 260 | -----KTTSAKETQSIEKTSAKDLAPTSKVLAKPTPKAET----- | 294 |
| Db | 706 | IRTFKESPKOILDPAARVTGMKKWPRTPKEAQSLE-----DLAGPKELFTQPSSESM | 760 |
| QY | 295 | -----TTK-----GPAITTPKE | 306 |
| Db | 761 | TDEKTTKIACKSPPPESVDPTSTKQWPKRSLRKADVEEFLALRLKLTSPSAGKAMLTTPK | 820 |
| QY | 307 | P-----TPT-----TPKEPAST----- | 318 |
| Db | 821 | AGGDEKDIKAFMGTPVQKLDLGLTFSGKQLOTPKEKAQALDLEDLAGFKELFQTPGHTEE | 880 |
| QY | 319 | -----TPKEP-----TPTTKSAPT-----TP | 335 |
| Db | 881 | LVAAGTKTKPCDSQSDPDVDTSTKQRPKRIRKADVEGELLACRLNMPKSAGKAMHTP | 940 |
| QY | 336 | KEPAP-----TTTSAPTTPKEPAPT----- | 356 |
| Db | 941 | K-----PSVGEKDIIFVGTVPQKLDLTENLTGSKRRFPQTPKEAQALEDLTGFKELFQTP | 997 |
| QY | 357 | -----TYKEPAPTTPKEPAPTTPKEPAPTTPKESAPTTPKE----- | 391 |
| Db | 998 | GHTEEAAGKTKTMCPESSPPESADT-----PTSTRQPKTFLEKRDVQKELSALKKL | 1051 |
| QY | 392 | -----PAPTTPKPAFTTPKEPA----- | 409 |
| Db | 1052 | TQTSGETHTDKVPGGEDKSINAFRETAKQKLDPAASVTSKGRHPT-KEKAQPLEDLA | 1110 |
| QY | 410 | -----PTTPK-----EPPT-TPKEPAPTKE----- | 430 |
| Db | 1111 | WKELFQTPVCTDKPTTHEKTKIACRSQDPDPTSTSSKQSKRSLRKVDVEEVEFFALRK | 1170 |
| QY | 431 | -----PAPT-----TPKEPA----- | 440 |
| Db | 1171 | RTPSAGKAMHTPKPAVSGEKNYAFMGTPVQKLDLTENLTGSKRRLQTPKEKAQLEDLA | 1230 |
| QY | 441 | -----PTAPKKA----- | 448 |
| Db | 1231 | GFKELFQTRGHTESMTNDKTAKVACKSSQDLDKNPASSKRRKLTSLGKGVKEELLAV | 1290 |
| QY | 449 | -----PTTPKPAFT-----TPKEPA----- | 464 |
| Db | 1291 | GKLTQTSGETHTTHEPTGDKGKMAKPMESPKIILDSASILTSGSKRQLRTPKGSVEPDD | 1350 |
| QY | 465 | -----PTTKESPTTPK-----EP-----APTTSKAP----- | 488 |
| Db | 1351 | LAGFIELFQTPSHTKB-SMTNEKTKVSYRASQDPLDVTPTSSKQPKRSLRKADTEEF | 1409 |
| QY | 489 | -----TTTKEPA----- | 495 |
| Db | 1410 | LAFKQTPSAGKAMHTPKPAVGEKODINTFLGTVPQKLDQGNLPGSNRRLQTRKEKAQ | 1469 |
| QY | 496 | -----PT-----TTK-----SAPTTPKEPSTTTKE----- | 516 |

| | | | |
|--|------|---|------|
| QY | 1087 | NOGIIII-----NPMU----- | 1096 |
| Db | 1077 | EQRIQMEQAKRLREBELLEKQLOEQEIEBKARNEMIQRKQMLQLEELKEAERQVILL | 1136 |
| QY | 1097 | -----SDETNICNGKVPDGLTTLRNGTLVAFRGHYFMWLSPPSPSPAR----- | 1140 |
| Db | 1137 | EOERLOEQERQRLIAEKEAIAFG-----SISTTTEASKPKYR-----LRPAQC | 1180 |
| QY | 1141 | -----RITEYWGIPSIDTTFVTRCNCE-GKTFFFKDSQYWRFTNDIKDAGYKPEIFKGFGG | 1195 |
| Db | 1181 | AAINKETRVNITDPSEWI--QKNCEFAKYF-----PEASCP----- | 1216 |
| QY | 1196 | LTSQIIVAALSTAKYKNWPESVYFF 1219 | |
| Db | 1217 | -----QIQALI-----ESCFAP 1228 | |
| RESULT 15 | | | |
| B48666 | | | |
| cell proliferation antigen Ki-67, short form - human | | | |
| C:Species: Homo sapiens (man) | | | |
| C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004 | | | |
| C:Accession: B48666 | | | |
| R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde | | | |
| J. Cell Biol. 123, 513-522, 1993 | | | |
| A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq | | | |
| ins. | | | |
| A:Reference number: A48666; MUID:94043435; PMID:8227122 | | | |
| A:Accession: B48666 | | | |
| A>Status: preliminary | | | |
| A:Molecule type: mRNA | | | |
| A:Residues: 1-2897 <SCH> | | | |
| A:Cross-references: EMBL:X65551 | | | |
| C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat | | | |
| F;29-91/Domain: kinase interaction domain homology <KIH> | | | |
| Query Match 12.0%; Score 881.4; DB 2; Length 2897; | | | |
| Best Local Similarity 14.1%; Pred. No. 3.2e-10; | | | |
| Matches 447; Conservative 151; Mismatches 457; Indels 2126; Gaps 111; | | | |
| QY | 3 | WKT-----LPYILLLSVFIQVSSQDLSSCAGRCGEGYSRDATNCND | 47 |
| Db | 2 | WPTRLVTIKRSGVDGHPFL-----SLSTCL----- | 28 |
| QY | 48 | YNCQHYMECCPDFKRVCTAELSCRCFESPERGECDDAQ-----CKKYDKCCPD | 99 |
| Db | 29 | -----FGRGIECDIRIQLPWVSKQHC----- | 50 |
| QY | 100 | YESFCAEVH-----NPTS----- | 112 |
| Db | 51 | -----IEIHEQEAHLNFSSTNPQVNGSVIDEVRLKHGDVITIDRSFYENESLONG | 105 |
| QY | 113 | -----PPS----- | 115 |
| Db | 106 | RKSTFPRKIREQPARVRSSPSSDDEKAQDSKAYSKITEGKVGSNPELFDENLPPN | 165 |
| QY | 116 | S-----KKAPPPSG----- | 124 |
| Db | 166 | TPLKRGAPTCKKSLVMHTPPVLKKIIEQOPQSGKQSGSEIHVEVKAQSLIVISPPAPS | 225 |
| QY | 125 | -----ASQTKSTTKRSPKPNKKTK----- | 146 |
| Db | 226 | PKRTPVASDQRRRSCKTAPASSSKSQTEVPKRGGERVATCLQKRVISIRSQHDILQMTCS | 285 |
| QY | 147 | -----KVIESBEITEVKNKKNRTKKP-TPKPPV----- | 175 |
| Db | 286 | KRRGASEANLIVAKSWADVVKLGAKQQTQVKVHKGPQRSNMKQRRRPAATPKKPVGEVHS | 345 |
| QY | 176 | -----VDEAG----- | 180 |
| Db | 346 | QFSTGHANSPTIILIGKAHTEKVVHVPARYVRLNNFISNQKMDFKEDLSLGAEMFKTPVK | 405 |

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 105.696 Seconds
(without alignments)
4449.477 Million cell updates/sec

Title: SEQ1-H

Perfect score: 7064

Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRSGQTLSKVWVNCV 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 7058 | 99.9 | 1311 | 8 | Adk67912 Human ext |
| 2 | 7044.7 | 99.7 | 1404 | 2 | Aar26049 MSF precu |
| 3 | 7044.7 | 99.7 | 1404 | 4 | Aab29773 Human meg |
| 4 | 7044.7 | 99.7 | 1404 | 4 | Aab60568 Human meg |
| 5 | 7044.7 | 99.7 | 1404 | 8 | Adm98014 Human meg |
| 6 | 7041.7 | 99.7 | 1415 | 4 | Aau32262 Novel hum |
| 7 | 7029.7 | 99.5 | 1404 | 7 | Adk65839 Angiogene |
| 8 | 6793.9 | 96.2 | 1270 | 8 | Adk67911 Human ext |
| 9 | 6769.9 | 95.8 | 1320 | 7 | Adk65819 Angiogene |
| 10 | 6468.7 | 91.6 | 1299 | 4 | Aam24322 Human ESF |
| 11 | 3552 | 50.3 | 902 | 4 | Aam29778 Human MSF |
| 12 | 2929 | 41.5 | 551 | 4 | Abu53253 Human tes |
| 13 | 2820 | 41.3 | 546 | 4 | Abu53252 Human tes |
| 14 | 2850.9 | 40.4 | 538 | 5 | Aao18834 3' cartil |
| 15 | 2757 | 39.0 | 513 | 4 | Abu53254 Human tes |
| 16 | 2009 | 28.7 | 472 | 4 | Aab60569 Bovine MS |
| 17 | 1958.9 | 27.7 | 452 | 2 | Aar80041 Human meg |
| 18 | 1398.2 | 19.8 | 5179 | 4 | Aam24516 C899p pre |
| 19 | 1398.2 | 19.8 | 5179 | 6 | Abp55365 Human col |
| 20 | 1398.2 | 19.8 | 5179 | 6 | Abc07258 Human p53 |
| 21 | 1398.2 | 19.8 | 5179 | 7 | Add48091 Human Pro |
| 22 | 1398.2 | 19.8 | 5179 | 7 | Add44998 Human Pro |
| 23 | 1307.1 | 18.5 | 292 | 5 | Aau11261 Human HAP |
| 24 | 1154.2 | 16.3 | 1664 | 2 | Aaw43106 C. thermo |
| 25 | 1139.7 | 16.1 | 8991 | 6 | Abu08487 S. pneumo |

RESULT 1

ADK67912
ID ADK67912 standard; protein; 1311 AA.

XX AC ADK67912;

DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMBS) polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;
antisthmatic; antiinflammatory; antidiabetic; neuroprotective;
muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;
virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
cytostatic; gene therapy.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers
FT Peptide 1..30

FT /label= Signal peptide
FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30
according to identification method"

FT WO2004013292-A2.

XX 12-FEB-2004.

XX 30-JUL-2003; 2003WO-US024084.

XX 02-AUG-2002; 2002US-0400810P.

XX 19-SEP-2002; 2002US-0412197P.

XX 04-OCT-2002; 2002US-0416004P.

XX 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

PI Richardson TW, Emerling BM, Lindquist EA, Ramkumar J;
PI Lee SY;

XX WPI; 2004-157116/15.

XX N-FSDB; ADK67917.

XX New extracellular messengers and nucleic acids, useful for diagnosing,
treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
autoimmune thyroiditis.

ADL23265 Human MUC
ABU53144 Human tes
ABU53255 Human tes
ABB69806 Drosophil
AGB38942 Arabidops
APB43908 MUC5B par
ABD71434 Drosophil
ABB61364 Drosophil
Aao18833 5' cartil
ABB58595 Drosophil
ABU53141 Human tes
ABU53141 Human tes
ABU53143 Human tes
ABB60403 Drosophil
ABU53155 Human tes
ADN39110 Cancer/an
ADI21202 Novel hum
ABB68397 Drosophil
ABU53150 Human tes
ABU53149 Human tes
ABU53151 Human tes

ALIGNMENTS

RESULT 1

ADK67912
ID ADK67912 standard; protein; 1311 AA.

XX AC ADK67912;

DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMBS) polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic;
antisthmatic; antiinflammatory; antidiabetic; neuroprotective;
muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;
virucide; fungicide; antiparasitic; protozoacide; antihelminthic;
cytostatic; gene therapy.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers
FT Peptide 1..30

FT /label= Signal peptide
FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30
according to identification method"

FT WO2004013292-A2.

XX 12-FEB-2004.

XX 30-JUL-2003; 2003WO-US024084.

XX 02-AUG-2002; 2002US-0400810P.

XX 19-SEP-2002; 2002US-0412197P.

XX 04-OCT-2002; 2002US-0416004P.

XX 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

XX Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

PI Richardson TW, Emerling BM, Lindquist EA, Ramkumar J;
PI Lee SY;

XX WPI; 2004-157116/15.

XX N-FSDB; ADK67917.

XX New extracellular messengers and nucleic acids, useful for diagnosing,
treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes
mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or
autoimmune thyroiditis.

XX PS Claim 60; SEQ ID NO 5; 165pp; English.

XX CC The present sequence is that of novel human extracellular messenger

CC (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology

CC to human megakaryocyte stimulating factor. The invention provides EXMES

CC polynucleotides and polypeptides, as well as expression vectors, host

CC cells, antibodies, agonists and antagonists, and methods for diagnosing,

CC treating or preventing disorders associated with aberrant expression of

CC EXMES, especially autoimmune and inflammatory disorders, cell

CC proliferative disorders and endocrine disorders. e.g. adult respiratory

CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's

CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,

CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,

CC parasitic, protozoal or helminthic infections, cancers, autoimmune

CC thyroiditis, cretinism, plummer's disease or thyroid carcinoma.

CC Embodiments also provide methods for using the purified EXMES and/or

CC their encoding polynucleotides for facilitating the drug discovery

CC process, including determining of efficacy, dosage, toxicity and

CC pharmacology, and for investigating the pathogenesis of diseases and

CC medical conditions.

XX SQ Sequence 1311 AA;

Query Match 99.9%; Score 7059; DB 8; Length 1311;

Best Local Similarity 99.9%; Pred. No. 3.4e-171;

Matches 1310; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCQDYNCHYMECCPDF 60

DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCQDYNCHYMECCPDF 60

QY 61 KRYCTAELSCGRCFESFERGREGDCDAQCKKYDKCCPDYESCAEVKDNKXRTKKKPT 120

DB 61 KRYCTAELSCGRCFESFERGREGDCDAQCKKYDKCCPDYESCAEVKDNKXRTKKKPT 120

QY 121 PKPVDVDEAGSLDNGDFKVTTPDTSITQHNKYSTSPKITTAKPINRPSLPNSDTSKE 180

DB 121 PKPVDVDEAGSLDNGDFKVTTPDTSITQHNKYSTSPKITTAKPINRPSLPNSDTSKE 180

QY 181 TSLTVNKETTVETKETTITNNKQTSIDGKEKTTSAKETQSIKTSKADLAFTSKVLAKPTP 240

DB 181 TSLTVNKETTVETKETTITNNKQTSIDGKEKTTSAKETQSIKTSKADLAFTSKVLAKPTP 240

QY 241 KAETTTKGPALTTTKEPTTTPKEPASTTPKEPPTTIKSAPTTPKEPAPTTTKSAPTTP 300

DB 241 KAETTTKGPALTTTKEPTTTPKEPASTTPKEPPTTIKSAPTTPKEPAPTTTKSAPTTP 300

QY 301 KEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAP 360

DB 301 KEPAPTTTKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAP 360

QY 361 TTPKEPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 420

DB 361 TTPKEPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 420

QY 421 PSPPTTPKEPAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTP 480

DB 421 PSPPTTPKEPAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTP 480

QY 481 KKPAPTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAP 540

DB 481 KKPAPTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAP 540

QY 541 PTTPEKPAPTTPPEELAPTTTPEEPTTTPPEEPAFTTPKAAAPNTTPKEPAPTTTPKEPAP 600

DB 541 PTTPEKPAPTTPPEELAPTTTPEEPTTTPPEEPAFTTPKAAAPNTTPKEPAPTTTPKEPAP 600

QY 601 KEPAPTTPKEPAPTTTPKGAFTTTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKPAFTT 660

DB 601 KEPAPTTPKEPAPTTTPKGAFTTTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKPAFTT 660

QY 661 PKGTAPTTTPKEPAPTTTPKGTAPTTTLKEPAPTTTPKPAKELAPTTTKEPTST 720

Db 661 PKGTAPTTTPKEPAPTTTPKGTAPTTTLKEPAPTTTPKPAKELAPTTTKEPTST 720

QY 721 TSDKPAPTTPKETAPTTPKEPAPTTTPKPAKELAPTTTKEPTSTSEVSTPTTTPKEPTTHKSP 780

Db 721 TSDKPAPTTPKETAPTTPKEPAPTTTPKPAKELAPTTTKEPTSTSEVSTPTTTPKEPTTHKSP 780

QY 781 DESTPELSAEPPTPKALENSPKFPGVPTTKTAAKPEMTTAAKDKTTERDLATTPETTTA 840

Db 781 DESTPELSAEPPTPKALENSPKFPGVPTTKTAAKPEMTTAAKDKTTERDLATTPETTTA 840

QY 841 AFKMTKETATTTTEKTESKITATTTQVSTTTTODTTPFKITTLKTTTLAPKVTTHKTTT 900

Db 841 AFKMTKETATTTTEKTESKITATTTQVSTTTTODTTPFKITTLKTTTLAPKVTTHKTTT 900

QY 901 TTEIMNKPBETAKPKDRATNSKATTPKPKOKPTKAPKKPTSTKKPKTMPVRVKPTTPTPR 960

Db 901 TTEIMNKPBETAKPKDRATNSKATTPKPKOKPTKAPKKPTSTKKPKTMPVRVKPTTPTPR 960

QY 961 KMTSTMPELNPTSRIAEAMLTQTTTRENQTPNSKLVENPKSEDAAGAGETPHMLLRPHV 1020

Db 961 KMTSTMPELNPTSRIAEAMLTQTTTRENQTPNSKLVENPKSEDAAGAGETPHMLLRPHV 1020

QY 1021 FMPEVTPDMDYLPVRVNOGIIINPMLSDETNI CNQKPVVDGLTTLRNGTLVAFRGHYFWML 1080

Db 1021 FMPEVTPDMDYLPVRVNOGIIINPMLSDETNI CNQKPVVDGLTTLRNGTLVAFRGHYFWML 1080

QY 1081 SPSPSPARRITEVWGIPIPSIDTTFTRCNCEKTFEFDKDSQYWRFTNDIKDAGYKPIF 1140

Db 1081 SPSPSPARRITEVWGIPIPSIDTTFTRCNCEKTFEFDKDSQYWRFTNDIKDAGYKPIF 1140

QY 1141 KGFGLTGTQIIVAAALSTAKYNWPESVYFFKRGSSIQQYIYKQEPVQKCPRRPALNYPVY 1200

Db 1141 KGFGLTGTQIIVAAALSTAKYNWPESVYFFKRGSSIQQYIYKQEPVQKCPRRPALNYPVY 1200

QY 1201 GEMTQVRRRFRFARAIGPSQHTTIRIQYSPARLAYQKGVLHNKVSILMRGLPNVVTSA 1260

Db 1201 GEMTQVRRRFRFARAIGPSQHTTIRIQYSPARLAYQKGVLHNKVSILMRGLPNVVTSA 1260

QY 1261 ISLPNIRKPDGYDYAFSKDOYVNIQVPSRTARAITTRSGQTLISKWYNCP 1311

Db 1261 ISLPNIRKPDGYDYAFSKDOYVNIQVPSRTARAITTRSGQTLISKWYNCP 1311

RESULT 2

AAR26049

ID AAR26049 standard; protein; 1404 AA.

XX AC AAR26049;

XX AC AC

XX DT 25-MAR-2003 (revised)

XX DT 02-FEB-1993 (first entry)

XX MSF precursor.

XX DE DE

XX KW KW

XX KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;

XX KW stability; proteolytic cleavage; adhesion; alternative splicing.

XX OS Synthetic.

XX OS OS

XX FH Key

XX FH Region

XX FT /label= Exon_I

XX FT 26..67

XX FT Region

XX FT /label= Exon_II

XX FT 67..107

XX FT Region

XX FT /label= Exon_III

XX FT 107..157

XX FT Region

XX FT /label= Exon_IV

XX FT 157..200

XX FT Region

XX FT /label= Exon_V

XX FT 200..1141

XX FT Region

XX FT /label= Exon_VI

| | | | | | | |
|----|---|---|-----|------|--|------|
| FT | Region | 1166..1212 | Db | 61 | KRVCTAELSCGRCFESFERGRCDCDAQCKYDKCCPDYBSFCAEVHNPTSPSPSKAP | 120 |
| FT | Region | /label= Exon_VIII | Qy | 107 | ----- | 106 |
| FT | Region | 1213..1266 | Db | 121 | PPSGASQTIKSTTKRSPKPNKKTKKVI | 180 |
| FT | Region | /label= Exon_IX | Qy | 107 | -----VKDNKNRTKKPPKPPVVD | 147 |
| FT | Region | 1266..1331 | Db | 181 | KIKSKNSAANRELQKKLVKDNKNRTKKKPT | 240 |
| FT | Region | /label= Exon_X | Qy | 148 | TOHNKVSTSPKITTAKPINPRPSLPNSDTS | 207 |
| FT | Region | 1331..1373 | Db | 241 | TOHNKVSTSPKITTAKPINPRPSLPNSDTS | 300 |
| FT | Region | /label= Exon_XI | Qy | 208 | KEKTSKETSQSIKTSKSLTNKETT | 267 |
| FT | Region | 1373..1404 | Db | 301 | KEKTSKETSQSIKTSKSLTNKETT | 360 |
| FT | Region | /label= Exon_XII | Qy | 268 | TTPKPTPTTIKSAPTTPKBPATTTKSA | 327 |
| FT | Region | 1411..1166 | Db | 361 | TTPKPTPTTIKSAPTTPKBPATTTKSA | 420 |
| FT | Region | /label= Exon_VII | Qy | 328 | APTTIKSAPTTPKBPATTTPKBPATTT | 387 |
| PN | WO9213075-A1. | | Db | 421 | APTTIKSAPTTPKBPATTTPKBPATTT | 480 |
| XX | 06-AUG-1992. | | Qy | 388 | EPAPTAPKAPATTTPKBPATTTPKBP | 447 |
| XX | 17-JAN-1992; 92WO-US000433. | | Db | 481 | EPAPTAPKAPATTTPKBPATTTPKBP | 540 |
| XX | 18-JAN-1991; 91US-00643502. | | Qy | 448 | TTKSAPTTPKBPATTTPKBPATTTPK | 507 |
| XX | 10-SEP-1991; 91US-00757022. | | Db | 541 | TTKSAPTTPKBPATTTPKBPATTTPK | 600 |
| XX | (GEM) GENETICS INST INC. | | Qy | 508 | APTAPKAPATTTPKBPATTTPKBP | 567 |
| XX | Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG; | | Db | 601 | APTAPKAPATTTPKBPATTTPKBP | 660 |
| XX | WPI; 1992-284660/34. | | Qy | 568 | PREPAPTTKAAANTPKBPATTTPKBP | 627 |
| XX | N-PSDB; AAQ27223. | | Db | 661 | PREPAPTTKAAANTPKBPATTTPKBP | 720 |
| XX | New human mega-karyocyte stimulating factors - for treating immune | | Qy | 628 | APTAPKAPATTTPKBPATTTPKBP | 687 |
| XX | deficiencis, cancer, exposure to radiation or drugs, bacterial and viral | | Db | 721 | APTAPKAPATTTPKBPATTTPKBP | 780 |
| XX | infections, etc. | | Qy | 688 | TAPTTLKBPATTTPKBPATTTPKBP | 747 |
| XX | Claim 1, 2 and 3; Fig 1; 87pp; English. | | Db | 781 | TAPTTLKBPATTTPKBPATTTPKBP | 840 |
| XX | The sequence given is a full length translation from the megakaryocyte | | Qy | 748 | KDAPTTPPTTSEVSTPTTKBPATTT | 807 |
| XX | stimulating factor (MSF) precursor. The sequence covered by exons II, III | | Db | 841 | KDAPTTPPTTSEVSTPTTKBPATTT | 900 |
| XX | and IV encodes megakaryocyte stimulating factor (MSF). This sequence is | | Qy | 808 | TKTAAATPEMTTAKDKTTRDLTBT | 867 |
| XX | modified by the addition of an N-terminal sequence encoding a secretory | | Db | 901 | TKTAAATPEMTTAKDKTTRDLTBT | 960 |
| XX | leader, an initiating methionine preceding exon II and a terminating | | Qy | 868 | TSTTTQDTPPKITTLTKTTLLAPKV | 927 |
| XX | codon following exon IV. The cDNA sequence given contains sequences | | Db | 961 | TSTTTQDTPPKITTLTKTTLLAPKV | 1020 |
| XX | derived from human megakaryocyte colony stimulating factor (meg-CSF). | | Qy | 928 | PQKPTKAPKPTSTTKKPMRVRKPKT | 987 |
| XX | Exon I contains the initiating methionine, and encodes a classical | | Db | 1021 | PQKPTKAPKPTSTTKKPMRVRKPKT | 1080 |
| XX | mammalian protein secretion signal sequence. The sequence encoding the | | Qy | 988 | QTPNSKLVVNPKSEDAGGAGETPHML | 1047 |
| XX | original meg-CSF includes exons II-IV and is thought to terminate in the | | Db | 1081 | QTPNSKLVVNPKSEDAGGAGETPHML | 1140 |
| XX | region between amino acid residues 134 - 147. The primary transcript of | | Qy | 1048 | DETNI CNKGPVDGLTTLRNGTLVAF | 1107 |
| XX | this gene may be cleaved in different ways to yield a family of mRNA's | | Db | 1141 | DETNI CNKGPVDGLTTLRNGTLVAF | 1200 |
| XX | each encoding a different MSF protein. Exons V and VI are thought to be | | | | | |
| XX | related to the activity of the factor and are also implicated in the | | | | | |
| XX | stability, folding and processing of the molecule. These exons are also | | | | | |
| XX | thought to play a role in the observed synergy of MSF with other | | | | | |
| XX | cytokines. Exons V - XII are believed to be implicated in the processing | | | | | |
| XX | or folding of the appropriate structure of the resulting factor, ie. one | | | | | |
| XX | or more of these exons may contain sequences which direct proteolytic | | | | | |
| XX | cleavage, adhesion, organisation of the cellular matrix or extracellular | | | | | |
| XX | matrix processing. Both naturally occurring and non-naturally occurring | | | | | |
| XX | MSF's may be characterised by various combinations of alternatively | | | | | |
| XX | spliced exons from this sequence, with the exons spliced together in | | | | | |
| XX | differing orders to form different members of the MSF family. (Updated on | | | | | |
| XX | 25-MAR-2003 to correct PN field.) | | | | | |
| XX | Sequence 1404 AA; | | | | | |
| XX | Query Match | 99.7%; Score 7044.7; DB 2; Length 1404; | | | | |
| XX | Best Local Similarity | 93.4%; Pred. No. 8.2e-171; | | | | |
| XX | Matches 1311; Conservative | 0; Mismatches | | | | |
| XX | | 0; Indels | | | | |
| XX | | 93; Gaps | | | | |
| XX | 1 MAWKTLPIYLLLLSVFVIQQVSSQDLS | CCAGRCGEGYSRATCNCDYNQHYMECCPDF | 60 | | | |
| XX | 1 MAWKTLPIYLLLLSVFVIQQVSSQDLS | CCAGRCGEGYSRATCNCDYNQHYMECCPDF | 60 | | | |
| XX | 61 KRVCTAELSCGRCFESFERGRCDCDAQCKYDKCCPDYBSFCAE | ----- | 106 | | | |

QY 1108 RCNCEGKTFKFDKSYWRFTNDIKDAGYKPKIFKGGGLTGQIVAAALSTAKYKNWPESY 1167
DB 1201 RCNCEGKTFKFDKSYWRFTNDIKDAGYKPKIFKGGGLTGQIVAAALSTAKYKNWPESY 1260
QY 1168 FFKRGSGTQYIYKQEPVQKCPGRPALNYPVYGEVQVRRRFRRAIGPSOHTIRIQY 1227
DB 1261 FFKRGSGTQYIYKQEPVQKCPGRPALNYPVYGEVQVRRRFRRAIGPSOHTIRIQY 1320
QY 1228 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1287
DB 1321 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1380
QY 1288 PSRTARAITRSGQTLKQWYNCP 1311
DB 1381 PSRTARAITRSGQTLKQWYNCP 1404

RESULT 3
AAB29773
ID AAB29773 standard; protein; 1404 AA.
XX AAB29773;
XX 28-FEB-2001 (first entry)
DT Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
XX Human MSF; megakaryocyte stimulating factor; tribonectin;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX Homo sapiens.
OS
XX WO200064930-A2.
PN 02-NOV-2000.
XX 24-APR-2000; 2000WO-US010953.
XX 23-APR-1999; 99US-00298970.
XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX Jay GD;
XX WPI; 2001-024673/03.
XX N-PSDB; AAC81498.
XX Novel tribonectin polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety.
XX
XX Claim 3; Page 7; 47pp; English.

XX The invention relates to a human tribonectin which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonectin has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSF-derived tribonectin; a biocompatible composition comprising a human
CC tribonectin for inhibiting tissue adhesion formation; and a method of
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonectin and DNA encoding it are useful in the treatment of
CC osteoarthritis, where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonectin,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal, where the injury is caused by a surgical insertion or trauma, or

CC an artificial device e.g., an orthopaedic implant. In particular, one of
CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
CC used in gene therapy. The present sequence represents human MSF
XX
SQ Sequence 1404 AA;
Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
Best Local Similarity 93.4%; Pred. No. 8.2e-171;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KVCCTAELSCKRCFESFERGECDCDAOCKKYDKCCPDYEFCAE- 106
DB 61 KVCCTAELSCKRCFESFERGECDCDAOCKKYDKCCPDYEFCAE- 120
QY 107 106
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESEBETEEHSVSENQESSSSSSSSSSTIW 180
QY 107 -----VKONKORTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 147
DB 181 KIKSSKNSAANRELQKLVKDKNKRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 148 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKETTNNKQSTSDG 207
DB 241 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKETTNNKQSTSDG 300
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKGPPTTPKPEPAS 267
DB 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKGPPTTPKPEPAS 360
QY 268 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEP 327
DB 361 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTTKPEPAPTTTKPEPAPTTTKPEP 420
QY 328 APTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTPK 387
DB 421 APTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTPK 480
QY 388 EPAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPSPTTKPEPAPTTTKSAPTTTKPEAPT 447
DB 481 EPAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPSPTTKPEPAPTTTKSAPTTTKPEAPT 540
QY 448 TTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKKRAPTTTKPEPAPTTTKPEAPT 507
DB 541 TTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKKRAPTTTKPEPAPTTTKPEAPT 600
QY 508 APAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEAPT 567
DB 601 APAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEAPT 660
QY 568 PEPAPPTPKAAAPNTKPEPAPTTTKPEPAPTTTKPEPAPTTTKETAPTTPKGTAPTTLKEP 627
DB 661 PEPAPPTPKAAAPNTKPEPAPTTTKPEPAPTTTKPEPAPTTTKETAPTTPKGTAPTTLKEP 720
QY 628 APPTPKAPKAPKELAPTTTKPEPTSTSDKAPTTTKGTAPTTPKGTAPTTPKPEAPTTPKG 687
DB 721 APPTPKAPKAPKELAPTTTKPEPTSTSDKAPTTTKGTAPTTPKGTAPTTPKPEAPTTPKG 780
QY 688 TAPPTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTKETAPTTPKGTAPTTPKPEAPTTPK 747
DB 781 TAPPTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTKETAPTTPKGTAPTTPKPEAPTTPK 840
QY 748 KPAPTTPETPTPTTSEVSTPTTKPEPTTIHKSPEDESTPELSAETPKALENSPKERGVT 807
DB 841 KPAPTTPETPTPTTSEVSTPTTKPEPTTIHKSPEDESTPELSAETPKALENSPKERGVT 900
QY 808 TKTPAATKPEMTTAKDKTERDURTPETTTAAKPKTKETATTTKTTESKITATTTQV 867
DB 901 TKTPAATKPEMTTAKDKTERDURTPETTTAAKPKTKETATTTKTTESKITATTTQV 960

QY 868 TSTTTQDTPPKITLTKTTTLPKVTTKTKITTTTEIMNKPEETAKPKDRATNSKATPK 927
 DB 961 TSTTTQDTPPKITLTKTTTLPKVTTKTKITTTTEIMNKPEETAKPKDRATNSKATPK 1020
 QY 928 POKETKAPKSTSTKPKTMRVRKPKTTPTRKMTSTMPNLNPTSRIAEAMLQTTTRPN 987
 DB 1021 POKETKAPKSTSTKPKTMRVRKPKTTPTRKMTSTMPNLNPTSRIAEAMLQTTTRPN 1080
 QY 988 QTPNSKLVEVNPKEDEAGGEGETHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1047
 DB 1081 QTPNSKLVEVNPKEDEAGGEGETHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1140
 QY 1048 DETNINCNGKPDVGLTLLRNGTLVAFRGHYFWMLSPFPSPARRITVWGPSPIDTVET 1107
 DB 1141 DETNINCNGKPDVGLTLLRNGTLVAFRGHYFWMLSPFPSPARRITVWGPSPIDTVET 1200
 QY 1108 RCNCEGKTFEFKDSQYWRFNIDKAGYKPIPKFGFGLTGQIVAAALSTAKYNWPEVY 1167
 DB 1201 RCNCEGKTFEFKDSQYWRFNIDKAGYKPIPKFGFGLTGQIVAAALSTAKYNWPEVY 1260
 QY 1168 FFKRGGSIOQYIKQBPVKCPGRRPALNYPVYGEVTOVRRRPFERRAIGPSQTHIRIQY 1227
 DB 1261 FFKRGGSIOQYIKQBPVKCPGRRPALNYPVYGEVTOVRRRPFERRAIGPSQTHIRIQY 1320
 QY 1228 SPARLAYQDKGVHLNHEVYKVSILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1287
 DB 1321 SPARLAYQDKGVHLNHEVYKVSILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1380
 QY 1288 PSRTARAITRRSQTLTKVWYNCP 1311
 DB 1381 PSRTARAITRRSQTLTKVWYNCP 1404

RESULT 4

AAB60568
 ID AAB60568 standard; protein; 1404 AA.
 XX
 AC AAB60568;
 XX
 DT 27-APR-2001 (first entry)
 XX
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 XX
 KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;
 XX antiarthritic.
 OS Homo sapiens.
 XX
 PN WO200107068-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 21-JUL-2000; 2000WO-US020002.
 XX
 PR 23-JUL-1999; 99US-0145328P.
 XX
 PR 19-JUL-2000; 2000US-00145328.
 XX
 PA (UYCA-) UNIV CASE WESTERN RESERVE.
 XX
 PI Warman ML;
 XX
 DR WPI; 2001-182721/18.
 XX
 PT New composition comprising the camptodactyly-arthropathy-coxa vara-
 PT pericarditis protein in combination with an anesthetic, useful for
 PT treating osteoarthritis, or as lubricants of tissue and joints.
 XX
 PS Example 1; Page; 34pp; English.
 XX
 CC The invention relates to a method of treating osteoarthritis via the

administration of a composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents human megakaryocyte stimulating factor (MSF, CACP protein). Note: This sequence is not given in its entirety in figure 4 of the specification, although a GenBank accession number was given. This sequence was therefore obtained from GenBank (U70316)

XX
 SQ Sequence 1404 AA;

Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
 Best Local Similarity 93.4%; Pred. No. 8.2e-171; Indels 93; Gaps 1;
 Matches 1311; Conservative 0; Mismatches 0;

QY 1 MAWKTIPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNQHYMECCPDF 60
 DB 1 MAWKTIPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNQHYMECCPDF 60
 QY 61 KRVTAEALCKGRCFESFERGECDCDAQCKYDKCCPDYESFCAE-----106
 DB 61 KRVTAEALCKGRCFESFERGECDCDAQCKYDKCCPDYESFCAE-----106
 QY 107 -----106
 DB 121 PPSGASQTIKSTIKRSPKPNKKTKKVIETSEHSVSESSSSSSSSSSSSSTIW 180
 QY 107 -----VKONKKNRTKKKTPKPPVVDEAGSLDNGDFKVTPTDTS 147
 DB 181 KIKSSKNSAANRELOKKLVKONKKNRTKKKTPKPPVVDEAGSLDNGDFKVTPTDTS 240
 QY 148 TQHNKVSTSPKLTAKPINPRESLPNSDTSKETSITVNETTETKTTNNKQSTDG 207
 DB 241 TQHNKVSTSPKLTAKPINPRESLPNSDTSKETSITVNETTETKTTNNKQSTDG 300
 QY 208 KEKTTSAKETQSIKTSKDLAPTSKVLAKPKAETTTKGPALTTTPKEPTTTPKEPAS 267
 DB 301 KEKTTSAKETQSIKTSKDLAPTSKVLAKPKAETTTKGPALTTTPKEPTTTPKEPAS 360
 QY 268 TTPKEPTTTPKSAPTTPKEPAPTTTKGAPTTTPKEPAPTTTPKEPAPTTTPKEP 327
 DB 361 TTPKEPTTTPKSAPTTPKEPAPTTTKGAPTTTPKEPAPTTTPKEPAPTTTPKEP 420
 QY 328 APTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 387
 DB 421 APTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 480
 QY 388 EPAPTAPKKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAP 447
 DB 481 EPAPTAPKKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAP 540
 QY 448 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKEPAPTTTPK 507
 DB 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTPKEPAPTTTPKEPAPTTTPK 600
 QY 508 APTAPKEPAPTTTPKETAPTTTPKLTTPPEKLAATTPPKAPAPTTPEELAPTTPE 567
 DB 601 APTAPKEPAPTTTPKETAPTTTPKLTTPPEKLAATTPPEKLAATTPPEELAPTTPE 660
 QY 568 PEEPAPTTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTATLKEP 627
 DB 661 PEEPAPTTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTATLKEP 720

```
QY 628 APTPKKAPKELAPTTTKETSTSDKPAPTTPKGTATTPKEPAPTTKPEAPTTKPG 687
DB 721 APTPKKAPKELAPTTTKETSTSDKPAPTTPKGTATTPKEPAPTTKPEAPTTKPG 780
QY 688 TAPTTLKPAPTTPKPAKELAPTTTKGTSTSDKPAPTTPKETAPTTPKEPAPTTK 747
DB 781 TAPTTLKPAPTTPKPAKELAPTTTKGTSTSDKPAPTTPKETAPTTPKEPAPTTK 840
QY 748 KPAPTTPEPTPTTSEVSTPTTKETPTTIHKSPDESTPELSAETTPKALENSPKFPGVPT 807
DB 841 KPAPTTPEPTPTTSEVSTPTTKETPTTIHKSPDESTPELSAETTPKALENSPKFPGVPT 900
QY 808 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAKPMYKETAATTEKTESKITATTQV 867
DB 901 TKTPAAKPEMTTAKDKTTERDLRTTPETTTAAKPMYKETAATTEKTESKITATTQV 960
QY 868 TSTTTQDTTPKEITTLTKTTLLAPKVTITTKTITTTIEMNKPEETAKPKDRATNSKATTEK 927
DB 961 TSTTTQDTTPKEITTLTKTTLLAPKVTITTKTITTTIEMNKPEETAKPKDRATNSKATTEK 1020
QY 928 POKPTKAPKKPTSTKKPKTMPRVKPKTTPTPRKMTSTMPELNPTSGRIAEAMLQTTTRN 987
DB 1021 POKPTKAPKKPTSTKKPKTMPRVKPKTTPTPRKMTSTMPELNPTSGRIAEAMLQTTTRN 1080
QY 988 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1047
DB 1081 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1140
QY 1048 DETNINCNGPVDGLTTLRNGTLVAFRGHYFWMLSPTSPSPARRITEVWGPISPIDTFT 1107
DB 1141 DETNINCNGPVDGLTTLRNGTLVAFRGHYFWMLSPTSPSPARRITEVWGPISPIDTFT 1200
QY 1108 RCNCEGKTFEFDKSOYWRFTNDIKDAGYKPIPKFGFGLTGQIVAAIYSTAKYKNWPEVY 1167
DB 1201 RCNCEGKTFEFDKSOYWRFTNDIKDAGYKPIPKFGFGLTGQIVAAIYSTAKYKNWPEVY 1260
QY 1168 FFRGGSIQYIYKQEPVQKCPGRPALNYPVGEWTOVRRRPERAIGPSQHTIRIQY 1227
DB 1261 FFRGGSIQYIYKQEPVQKCPGRPALNYPVGEWTOVRRRPERAIGPSQHTIRIQY 1320
QY 1228 SPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKPDCGYDYAFSKQYYNIDV 1287
DB 1321 SPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKPDCGYDYAFSKQYYNIDV 1380
QY 1288 PSRTARAITRSGOTLSKWYNCP 1311
DB 1381 PSRTARAITRSGOTLSKWYNCP 1404

RESULT 5
ADM98014
ID ADM98014 standard; protein; 1404 AA.
XX AC ADM98014;
XX
XX 01-JUL-2004 (first entry)
XX Human megakaryocyte stimulating factor (MSF).
XX lubricating polypeptide; O-linked oligosaccharide; joint lubrication;
XX CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;
XX megakaryocyte stimulating factor; MSF.
XX
XX Homo sapiens.
XX
XX US2004072741-A1.
XX
XX 15-APR-2004.
XX
XX 02-JUL-2001; 2001US-00897188.
XX
XX 23-APR-1999; 99US-00298970.
XX
XX 24-APR-2000; 2000US-00556246.
XX
```

```
XX (JAYG/) JAY G D.
XX Jay GD;
XX WPI; 2004-373948/35.
XX N-PSDB; ADM98015.
XX New tribonectin polypeptides and polynucleotides for lubricating joints
XX PT or other tissues to prevent or treat Camptodactyl-arthropathy-
XX PT pericarditis syndrome or osteoarthritis.
XX
XX Claim 1; SEQ ID NO 1; 34pp; English.
XX
XX The invention relates to a lubricating polypeptide and at least one O-
XX linked oligosaccharide. The composition and methods are useful for
XX lubricating joints or other tissues to prevent or treat camptodactyl-
XX arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present
XX sequence represents the amino acid sequence of the human megakaryocyte
XX stimulating factor (MSF).
XX
XX Sequence 1404 AA;
SQ
Query Match 99.7%; Score 7044.7; DB 8; Length 1404;
Best Local Similarity 93.4%; Pred. No. 8.2e-171;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCYDNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYGRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVTAEELCKGRCFESFERGECDCDAQCKKYDKCPCDYESFCAE----- 106
DB 61 KRVTAEELCKGRCFESFERGECDCDAQCKKYDKCPCDYESFCAEVHNPTSPSSKKAP 120
QY 107 ----- 106
DB 121 PPSGASQTIKSTTRGPKPPNKKTKKVIIESEITBEHSVSENGQSSSSSSSSSSSTIW 180
QY 107 -----VKONKKNRTKKKPKPPVVDVDEAGSGLDNGFKVTTPTDTST 147
DB 181 KIKSSKNNSAANRELQKKLVKONKKNRTKKKPKPPVVDVDEAGSGLDNGFKVTTPTDTST 240
QY 148 TQNKVSTSPKITTAKPINRPSLPPNSDTSKETSLTVNKKETTVEKTTTNNKQSTDG 207
DB 241 TQNKVSTSPKITTAKPINRPSLPPNSDTSKETSLTVNKKETTVEKTTTNNKQSTDG 300
QY 208 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTKGPALTTPEKPTTPKPEPAS 267
DB 301 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTKGPALTTPEKPTTPKPEPAS 360
QY 268 TTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKEP 327
DB 361 TTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKEP 420
QY 328 APITTKSAPTTTPKEPAPTTTPKPAPTTPKPEAPTTTPKEPTTPKPEAPTTTPK 387
DB 421 APITTKSAPTTTPKEPAPTTTPKPAPTTPKPEAPTTTPKEPTTPKPEAPTTTPK 480
QY 388 EPAPTAAPKPAPTTPKPEAPTTTPKPEAPTTTKESPTTPKPEAPTTTKSAPTTTKEPAPT 447
DB 481 EPAPTAAPKPAPTTPKPEAPTTTPKPEAPTTTKESPTTPKPEAPTTTKSAPTTTKEPAPT 540
QY 448 TTKSAPTTTPKEPSPTTTPKPEAPTTTPKPAPTTPKPAPTTPKPEAPTTTPKPEAPTTTCKP 507
DB 541 TTKSAPTTTPKEPSPTTTPKPEAPTTTPKPAPTTPKPAPTTPKPEAPTTTPKPEAPTTTCKP 600
QY 508 APTAPKPEAPTTTPKETAFTTPKKLTPTTPEKLAFTTPEKAPTTPEELAPTTPEEPTPTT 567
DB 601 APTAPKPEAPTTTPKETAFTTPKKLTPTTPEKLAFTTPEKAPTTPEELAPTTPEEPTPTT 660
QY 568 PEEPAPTTPKAAAPNTPKPEAPTTTPKPEAPTTTPKETAFTTPKETAFTTPKGTATTPK 627
DB
```


| | | | |
|----------|----------|--|------|
| QY | 448 | TTKSAPTTPKSPSPPTTTKEPAPTTPKKPAPTTPKKPAPTTPKKPAPTTPKKPAPTTTKP | 507 |
| Db | 552 | TTKSAPTTPKSPSPPTTTKEPAPTTPKKPAPTTPKKPAPTTPKKPAPTTTKP | 611 |
| QY | 508 | APTAPKEAPPTPKETATTTKLTPTTPEKLAFTTPEKAPTTPEELAPTTPEPTPT | 567 |
| Db | 612 | APTAPKEAPPTPKETATTTKLTPTTPEKLAFTTPEKAPTTPEELAPTTPEPTPT | 671 |
| QY | 568 | PEEPAPTTPKAAAPNTPKEPAPTTPKKPAPTTPKKPAPTTPKETAPTTPKGAPTTLKEP | 627 |
| Db | 672 | PEEPAPTTPKAAAPNTPKEPAPTTPKKPAPTTPKKPAPTTPKETAPTTPKGAPTTLKEP | 731 |
| QY | 628 | APTTPKPKAPKELAPTTTKEPTSTSDKAPPTTPKGTAPTTPKKPAPTTPKKPAPTTKPG | 687 |
| Db | 732 | APTTPKPKAPKELAPTTTKEPTSTSDKAPPTTPKGTAPTTPKKPAPTTPKKPAPTTKPG | 791 |
| QY | 688 | TAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTSDKAPPTTPKGTAPTTPKKPAPTTK | 747 |
| Db | 792 | TAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTSDKAPPTTPKGTAPTTPKKPAPTTK | 851 |
| QY | 748 | KPAPTTPEPTTPEPTTSEVSTPTTKEPTTTHKSPDESTPELSAETTPKALENSPKPGVPT | 807 |
| Db | 852 | KPAPTTPEPTTPEPTTSEVSTPTTKEPTTTHKSPDESTPELSAETTPKALENSPKPGVPT | 911 |
| QY | 808 | TXTPAATKPEMTTAKDKTTERDLRTPPETTTAAPKMTKETAATTTKTESKITATTQV | 867 |
| Db | 912 | TXTPAATKPEMTTAKDKTTERDLRTPPETTTAAPKMTKETAATTTKTESKITATTQV | 971 |
| QY | 868 | TSTTTQDTTTPFKITLTKITLAPKVTTKTITTTIEMNKPEBTAKPKDRATNSKATTPK | 927 |
| Db | 972 | TSTTTQDTTTPFKITLTKITLAPKVTTKTITTTIEMNKPEBTAKPKDRATNSKATTPK | 1031 |
| QY | 928 | POKPTKAPKPTSTKPKTWPVRKPKTTPTRKMTSTWPELNPTSRIAEAMLTQTTTRN | 987 |
| Db | 1032 | POKPTKAPKPTSTKPKTWPVRKPKTTPTRKMTSTWPELNPTSRIAEAMLTQTTTRN | 1091 |
| QY | 988 | QTPNSKLVEVNPKESDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS | 1047 |
| Db | 1092 | QTPNSKLVEVNPKESDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS | 1151 |
| QY | 1048 | DETNIENGKPEVDGLTTLRNGTLVAFRGHYFMMLSPPSPSPARRITEVWGIPSPIDTVFT | 1107 |
| Db | 1152 | DETNIENGKPEVDGLTTLRNGTLVAFRGHYFMMLSPPSPSPARRITEVWGIPSPIDTVFT | 1211 |
| QY | 1108 | RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYXNWPBSVY | 1167 |
| Db | 1212 | RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYXNWPBSVY | 1271 |
| QY | 1168 | FFKRGSGSIQYIYKQBPVQKCPGRRPALNVPVYGEVMTQVRRRPERAIGPSQTHIRIQY | 1227 |
| Db | 1272 | FFKRGSGSIQYIYKQBPVQKCPGRRPALNVPVYGEVMTQVRRRPERAIGPSQTHIRIQY | 1331 |
| QY | 1228 | SPARLAYQDGVHLNVEKVSILRGLPNVNTSATSLENIRKPDGYDYAFSKOQYNNIDV | 1287 |
| Db | 1332 | SPARLAYQDGVHLNVEKVSILRGLPNVNTSATSLENIRKPDGYDYAFSKOQYNNIDV | 1391 |
| QY | 1288 | PSRTARAITTRSGQTLKSVWYNCP | 1311 |
| Db | 1392 | PSRTARAITTRSGQTLKSVWYNCP | 1415 |
| RESULT 7 | | | |
| ID | ADK65839 | standard; protein; 1404 AA. | |
| XX | AC | ADK65839; | |
| XX | DT | 06-MAY-2004 (first entry) | |
| DE | XX | Angiogenesis-differentially expressed protein #53. | |
| XX | XX | cytostatic; cardiant; vasotropic; antiarteriosclerotic; | |
| KW | XX | | |

angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
gene expression; cancer; coronary artery disease; myocardial ischemia;
coronary arteriosclerosis; forensic medicine.

Homo sapiens.

WO2003066831-A2.

14-AUG-2003.

07-FEB-2003; 2003WO-US003848.

07-FEB-2002; 2002US-00067482.

10-JUN-2002; 2002US-00164595.

16-AUG-2002; 2002US-0403649P.

03-JAN-2003; 2003US-0437746P.

(ORIG-) ORIGENE TECHNOLOGIES INC.

Sun Z, Li X, Kovacs KF, Fan W, Jay G;

WPI; 2003-731502/69.

Determining the angiogenic index of a tissue or cell sample using
expression levels of differentially expressed genes, useful for
diagnosing or treating cancer, coronary artery disease, myocardial
ischemia and/or arteriosclerosis.

Disclosure; SEQ ID NO 78; 296pp; English.

The invention relates to a method of determining the angiogenic index of
a tissue or cell sample comprising assessing, in a sample, the expression
levels of one or more differentially-expressed gene from any of 34 DNA
sequences, given in the specification, where the levels are indicative of
the angiogenic index. The methods and compositions of the present
invention are useful for diagnosing, preventing and/or treating cancer,
coronary artery disease, myocardial ischemia or coronary
arteriosclerosis. They can also be used in research, drug discovery and
forensic medicine involving angiogenesis. This sequence corresponds to
one of the differentially expressed proteins of the invention.

Seq
Sequence 1404 AA;

Query Match 99.5%; Score 7029.7; DB 7; Length 1404;
Best Local Similarity 93.2%; Pred. No. 2e-170;
Matches 1308; Conservative 0; Mismatches 3; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSVFIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

QY 61 KRVTAEISCKRCFESFERGRCDDAOCKKYDKCCPDYESFCAE----- 106
Db 61 KRVTAEISCKRCFESFERGRCDDAOCKKYDKCCPDYESFCAEVHNTPSPSSKAP 120

QY 107 ----- 106

Db 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSSSTIR 180

QY 107 -----VKDNKKNTKKKPKPVVDVDEAGSLNGDFKVTTPDTST 147

Db 181 KIKSSKNSAANRELQKKLVKDNKKNTKKKPKPVVDVDEAGSLNGDFKVTTPDTST 240

QY 148 TQHNKVTSPKITTAKPINRPSLPNNSDTSKETSLSLVNKETTVETKTTTNKQTSIDG 207

Db 241 TQHNKVTSPKITTAKPINRPSLPNNSDTSKETSLSLVNKETTVETKTTTNKQTSIDG 300

QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPAETTKGPAITTKGPAITTKGPA 267

Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPAETTKGPAITTKGPAITTKGPA 360

QY 268 TTPKEPTTITKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPE 327

| | | | |
|----|------|--|------|
| Db | 361 | TTPEPTPTTKSAPTTKBPAPTTTKSAPTTKBPAPTTTKBPAPTTTKBPAPTTTKBP | 420 |
| Qy | 328 | APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 387 |
| Db | 421 | APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 480 |
| Qy | 388 | EPARTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 447 |
| Db | 481 | EPARTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 540 |
| Qy | 448 | TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 507 |
| Db | 541 | TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 600 |
| Qy | 508 | APTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 567 |
| Db | 601 | APTAPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 660 |
| Qy | 568 | PEEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 627 |
| Db | 661 | PEEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 720 |
| Qy | 628 | APTTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 687 |
| Db | 721 | APTTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 780 |
| Qy | 688 | TAPTTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 747 |
| Db | 781 | TAPTTLKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP | 840 |
| Qy | 748 | KBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 807 |
| Db | 841 | KBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 900 |
| Qy | 808 | TKTPAATKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 867 |
| Db | 901 | TKTPAATKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP | 960 |
| Qy | 868 | TSTTTQDTPPKITLTKTTLAPKVTTKITLTKTTLTKTTLTKTTLTKTTLTKTT | 927 |
| Db | 961 | TSTTTQDTPPKITLTKTTLAPKVTTKITLTKTTLTKTTLTKTTLTKTTLTKTT | 1020 |
| Qy | 928 | POKTPAKPKPTSTKPKTPMPKPKPTPTPKMTSTMPKMTSTMPKMTSTMPKMT | 987 |
| Db | 1021 | POKTPAKPKPTSTKPKTPMPKPKPTPTPKMTSTMPKMTSTMPKMTSTMPKMT | 1080 |
| Qy | 988 | OTPNKSLVEVNPKSDAGAGETPHMLLRPHVFNPEVTPDMVLPVFNQGIINPMLS | 1047 |
| Db | 1081 | OTPNKSLVEVNPKSDAGAGETPHMLLRPHVFNPEVTPDMVLPVFNQGIINPMLS | 1140 |
| Qy | 1048 | DETNICNGKPDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTVFT | 1107 |
| Db | 1141 | DETNICNGKPDGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPIDTVFT | 1200 |
| Qy | 1108 | RCNCEGKTFPKDSOYWRFTNDIKDAGYKPKIFKGFGLTGQIVAAALSTAKYKNWPSVY | 1167 |
| Db | 1201 | RCNCEGKTFPKDSOYWRFTNDIKDAGYKPKIFKGFGLTGQIVAAALSTAKYKNWPSVY | 1260 |
| Qy | 1168 | FFKRGGSIQOYIYKQEPVQKCPGRPALNYPVYGMTOVRRRRFERRAIGPSQTHIRIQY | 1227 |
| Db | 1261 | FFKRGGSIQOYIYKQEPVQKCPGRPALNYPVYGMTOVRRRRFERRAIGPSQTHIRIQY | 1320 |
| Qy | 1228 | SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPRIKPDGDYVAFSKQOYINIDV | 1287 |
| Db | 1321 | SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPRIKPDGDYVAFSKQOYINIDV | 1380 |
| Qy | 1288 | PSRTARAITTRSGQTLKSKWYNCP 1311 | |
| Db | 1381 | PSRTARAITTRSGQTLKSKWYNCP 1404 | |

RESULT 8
ADK67911

| | |
|----|--|
| ID | ADK67911 standard; protein; 1270 AA. |
| XX | ADK67911; |
| AC | 06-MAY-2004 (first entry) |
| XX | Human extracellular messenger (EXMES) polypeptide. |
| DE | Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic; |
| XX | antisthmatic; antiinflammatory; antidiabetic; neuroprotective; |
| KW | muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic; |
| KW | virucide; fungicide; antiparasitic; protozoacide; antihelminthic; |
| KW | cytostatic; gene therapy. |
| XX | Homo sapiens. |
| OS | |
| XX | |
| FH | Key |
| FT | Peptide |
| FT | 1. .29 |
| FT | /label= Signal peptide |
| FT | /notes="Spans residues 1 to 18, 20, 21, 24, 27 or 29 |
| FT | according to identification method" |
| XX | W02004013292-A2. |
| PN | |
| XX | 12-FEB-2004. |
| PD | |
| XX | 30-JUL-2003; 2003WO-US024084. |
| PF | |
| XX | 02-AUG-2002; 2002US-0400810P. |
| PR | 19-SEP-2002; 2002US-0412197P. |
| PR | 04-OCT-2002; 2002US-0416004P. |
| PR | 08-NOV-2002; 2002US-0424862P. |
| XX | (INCY-) INCYTE CORP. |
| PA | |
| XX | Elliot VS, Khare R, Tran UK, Swarnakar A, Marquis JP; |
| PI | Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J; |
| PI | Lee SY; |
| XX | WPI; 2004-157116/15. |
| DR | N-PSDB; ADK67916. |
| XX | New extracellular messengers and nucleic acids, useful for diagnosing, |
| PT | treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes |
| PT | mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or |
| PT | autoimmune thyroiditis. |
| XX | Claim 59; SEQ ID NO 4; 165pp; English. |
| PS | |
| XX | The present sequence is that of novel human extracellular messenger |
| CC | (EXMES) incyte ID NO: 7513017CD1 polypeptide. The protein shows homology |
| CC | to human megakaryocyte stimulating factor. The invention provides EXMES |
| CC | polynucleotides and polypeptides, as well as expression vectors, host |
| CC | cells, antibodies, agonists and antagonists, and methods for diagnosing, |
| CC | treating or preventing disorders associated with aberrant expression of |
| CC | EXMES, especially autoimmune and inflammatory disorders, cell |
| CC | proliferative disorders and endocrine disorders, e.g. adult respiratory |
| CC | distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's |
| CC | disease, diabetes mellitus, myasthenia gravis, osteoarthritis, Crohn's |
| CC | osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, |
| CC | parasitic, protozoal or helminthic infections, cancers, autoimmune |
| CC | thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. |
| CC | Embodiments also provide methods for using the purified EXMES and/or |
| CC | their encoding polynucleotides for facilitating the drug discovery |
| CC | process, including determining of efficacy, dosage, toxicity and |
| CC | pharmacology, and for investigating the pathogenesis of diseases and |
| CC | medical conditions. |
| XX | |
| SQ | Sequence 1270 AA; |

Query Match 96.2%; Score 6793.9; DB 8; Length 1270;
Best Local Similarity 96.8%; Pred. No. 1.7e-164;
Matches 1269; Conservative 0; Mismatches 1; Indels 41; Gaps 1;

| | | | |
|----|------|---|------|
| QY | 1 | MAWKTLPYLLILLLLSVFVIOQVSSODLSSCAGRCGEGYSRATCNCDYNCQHMECCPDF | 60 |
| DB | 1 | MAWKTLPYLLILLLLSVFVIOQVSSQ----- | 25 |
| QY | 61 | KRVCTAELSCGRCPESFERGECDCDAQCKYDKCCPDYESFCAEVKDNKNRTRKKKPT | 120 |
| DB | 26 | -----EUSCKGRCPESFERGECDCDAQCKYDKCCPDYESFCAEVKDNKNRTRKKKPT | 79 |
| QY | 121 | PKPPVVVDAGSGLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKPINRPSLPNSDTSKE | 180 |
| DB | 80 | PKPPVVVDAGSGLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKPINRPSLPNSDTSKE | 139 |
| QY | 181 | TSLTVNKETTVTKETTTINKOTSDGKEKTTSAKETQSIEKTSAKDIAPTSKVLAKPTP | 240 |
| DB | 140 | TSLTVNKETTVTKETTTINKOTSDGKEKTTSAKETQSIEKTSAKDIAPTSKVLAKPTP | 199 |
| QY | 241 | KAETTTKGPALTPKPEPTTPKPEASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTP | 300 |
| DB | 200 | KAETTTKGPALTPKPEPTTPKPEASTTPKPEPTPTTIKSAPTTPKPEPAPTTTKSAPTTP | 259 |
| QY | 301 | KEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 360 |
| DB | 260 | KEPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 319 |
| QY | 361 | TTPKGPPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 420 |
| DB | 320 | TTPKGPPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 379 |
| QY | 421 | PSPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 480 |
| DB | 380 | PSPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 439 |
| QY | 481 | KKPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 540 |
| DB | 440 | KKPAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 499 |
| QY | 541 | PTTPEKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 600 |
| DB | 500 | PTTPEKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 559 |
| QY | 601 | KPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 660 |
| DB | 560 | KPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 619 |
| QY | 661 | PKGTAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 720 |
| DB | 620 | PKGTAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 679 |
| QY | 721 | TSDKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 780 |
| DB | 680 | TSDKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTP | 739 |
| QY | 781 | DESTPELSAEPTEKALENSKEPGVPTTKPAATKPEMTTAKDKITPERDLRTTPETTA | 840 |
| DB | 740 | DESTPELSAEPTEKALENSKEPGVPTTKPAATKPEMTTAKDKITPERDLRTTPETTA | 799 |
| QY | 841 | APKMTKETATTTETKTTESKITATTTQVSTSTTQDTPPKITLTKTTTLAPKVTITTKITIT | 900 |
| DB | 800 | APKMTKETATTTETKTTESKITATTTQVSTSTTQDTPPKITLTKTTTLAPKVTITTKITIT | 859 |
| QY | 901 | TTTBMKPEATAKPKORATNSKATTPKPKPTKAPKPTSTKPKPTMPVRVKPKTTPTPR | 960 |
| DB | 860 | TTTBMKPEATAKPKORATNSKATTPKPKPTKAPKPTSTKPKPTMPVRVKPKTTPTPR | 919 |
| QY | 961 | KMTSTPELNPTSRIAEAMLTQTTTRPNQTPNSKLIVEVNPXSEDAGGAEGETPHMLLPHV | 1020 |
| DB | 920 | KMTSTPELNPTSRIAEAMLTQTTTRPNQTPNSKLIVEVNPXSEDAGGAEGETPHMLLPHV | 979 |
| QY | 1021 | FMPVETPDMOYLPRVFNQGGIIINPMLSDETNI CNKQVDGLTTLNGLTVAFRGHYFWML | 1080 |
| DB | 980 | FMPVETPDMOYLPRVFNQGGIIINPMLSDETNI CNKQVDGLTTLNGLTVAFRGHYFWML | 1039 |

Query Match

95.8%; Score 6769.9; DB 7; Length 1320;

Best Local Similarity 93.1%; Pred. No. 7.3e-164;
Matches 1267; Conservative 0; Mismatches 3; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLILLLSLVFVIOQVSSQDLSSCAGRCGEGYSDRATCNCYNCOHYMECCPDF 60
Db 1 MAWKTLPIYLILLLSLVFVIOQVSSQ----- 25
QY 61 KRVCIAELSKGRCFESFERGREGCDCAQCKYDKCCPDYESFCA----- 105
Db 26 -----ELSKGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEYVHNTPSPSSKKAP 79
QY 106 -----EVKONKGNRTKKKPTPKPPVDEAG 130
Db 80 PPSGASOTIKSTTKRSPKPNKKTKKVBSEITEVKONKGNRTKKKPTPKPPVDEAG 139
QY 131 SGLDNGDFKVTTPDTSITQHNKVSSTPKITITAKDINPRPSLPPNSDTSKETSITVKNKETT 190
Db 140 SGLDNGDFKVTTPDTSITQHNKVSSTPKITITAKDINPRPSLPPNSDTSKETSITVKNKETT 199
QY 191 VETKETTITNKQSTDKGKXTTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTTKGPA 250
Db 200 VETKETTITNKQSTDKGKXTTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTTKGPA 259
QY 251 LTTPEKPTPTPKEPASTTTKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKE 310
Db 260 LTTPEKPTPTPKEPASTTTKEPTPTTIKSAPTTPKEPAPTTTKSAPTTTKEPAPTTTKE 319
QY 311 PAPTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTPKKAPTTTPKEPAPTTTKEPTPTT 370
Db 320 PAPTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTPKKAPTTTPKEPAPTTTKEPTPTT 379
QY 371 PKSEPAPTTPKEPAPTTPKAPPAAPTTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAP 430
Db 380 PKSEPAPTTPKEPAPTTPKAPPAAPTTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAP 439
QY 431 PTTTKSAPTTTKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTPKKAPPAAPTTTKEPAP 490
Db 440 PTTTKSAPTTTKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTPKKAPPAAPTTTKEPAP 499
QY 491 PAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAP 550
Db 500 PAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAP 559
QY 551 TPEELAPTTPEEPTPTTPEEAPTTTPKAAAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 610
Db 560 TPEELAPTTPEEPTPTTPEEAPTTTPKAAAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 619
QY 611 TAPTTPKGTAPTLKEPAPTTTPKAPKAPKELAPTTTKEPSTTSDKAPTTTPKGTAPTTPK 670
Db 620 TAPTTPKGTAPTLKEPAPTTTPKAPKAPKELAPTTTKEPSTTSDKAPTTTPKGTAPTTPK 679
QY 671 EPAPTTPKEPAPTTTPKGTAPTLKEPAPTTTPKAPKAPKELAPTTTKEPSTTSDKAPTTTP 730
Db 680 EPAPTTPKEPAPTTTPKGTAPTLKEPAPTTTPKAPKAPKELAPTTTKEPSTTSDKAPTTTP 739
QY 731 KETAPTTPKEPAPTTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 790
Db 740 KETAPTTPKEPAPTTTPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 799
QY 791 PTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKITTERDLTTPPETTTAAAPKMTKETAT 850
Db 800 PTPKALENSPKEPGVPTTKTPAATKPEMTTAKDKITTERDLTTPPETTTAAAPKMTKETAT 859
QY 851 TTEKTTESKITATTQVSTTTQDTPPKITTLTKTTTLAPKVTTTKKITTITTEIMNKPEE 910
Db 860 TTEKTTESKITATTQVSTTTQDTPPKITTLTKTTTLAPKVTTTKKITTITTEIMNKPEE 919
QY 911 TAKPKDRATNSKATTPKQKPTKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 970
Db 920 TAKPKDRATNSKATTPKQKPTKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 979
QY 971 PTSRIAEMLQTTTRPNQTPNSKLVENPKSDEAGAGETPHMLLRPHVFMPEVTPDMD 1030

Db 980 PTSRIAEMLQTTTRPNQTPNSKLVENPKSDEAGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1031 YLPRVNPQGIINPMLSDETNICNGKPVVDGLTTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1090
Db 1040 YLPRVNPQGIINPMLSDETNICNGKPVVDGLTTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1099
QY 1091 RITEVWGIPIPIDIVFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI 1150
Db 1100 RITEVWGIPIPIDIVFTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQI 1159
QY 1151 VAALSTAKYKNWSPESVYFFKRGSGIQYIYKQBPVQKCPGRRPALNPVYVGMTOVERRR 1210
Db 1160 VAALSTAKYKNWSPESVYFFKRGSGIQYIYKQBPVQKCPGRRPALNPVYVGMTOVERRR 1219
QY 1211 FERAIGSQTHTIRIQYSPARLAYQDKGVLHNEVKVSIWRGLPNVVVTSALSINIRKPD 1270
Db 1220 FERAIGSQTHTIRIQYSPARLAYQDKGVLHNEVKVSIWRGLPNVVVTSALSINIRKPD 1279
QY 1271 GYDYAFSKDQYVYNDVPSRTARAITTRSQTLTSKWYNCP 1311
Db 1280 GYDYAFSKDQYVYNDVPSRTARAITTRSQTLTSKWYNCP 1320
RESULT 10
AAM24322
ID AAM24322 standard; protein; 1299 AA.
XX
AC AAM24322;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1847.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;
KW gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002687.
XX
PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX
(HYSE-) HYSEQ INC.
PA
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Wehrman T;
XX
DR WPI; 2001-476164/51.
DR N-PSDB; AAH98981.
XX
Isolated polypeptide for treatment of diseases, diagnostics, raising
antibodies and research use.
PS Claim 20; Page 1198-1201; 1275pp; English.
XX
The present invention provides the protein and coding sequences of novel
proteins from a variety of organisms, including human, dog, cat, horse,
cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
urchin and tomato. These were derived from expressed sequence tags (ESTs)
from the organism of interest. They can be used in diagnostics,
forensics, gene mapping, identification of mutations, to assess
biodiversity and for nutritional purposes. The present sequence is a
protein of the invention
XX

| | | |
|----|--|---|
| SQ | Sequence 1299 AA; | |
| | Query Match | 91.6%; Score 6468.7; DB 4; Length 1299; |
| | Best Local Similarity | 92.8%; Pred. No. 3.3e-156; |
| | Matches 1202; Conservative | 0; Mismatches 0; Indels 93; Gaps 1; |
| QY | 1 MAWKTLPIYLLLLLVFVJQQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF | 60 |
| DB | 1 MAWKTLPIYLLLLLVFVJQQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF | 60 |
| QY | 61 KEVCTAELSCKRCFESFERGECDCDAOCKYDKCCPDYEFCAE----- | 106 |
| DB | 61 KRVCFAELSCKRCFESFERGECDCDAOCKYDKCCPDYEFCAE----- | 106 |
| QY | 107 ----- | 106 |
| DB | 121 PPSGASQTIKSTKRSPPKPKKTKKVIESEBIEEHSVSENEQSSSSSSSSSTIW | 180 |
| QY | 107 -----VKONKNRTYKKPTPKPPVVDGAGSLDNGDFKVTTPDTST | 147 |
| DB | 181 KIKSSKNSAANRELQKLVKONKNRTYKKPTPKPPVVDGAGSLDNGDFKVTTPDTST | 240 |
| QY | 148 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKEITTTNKQSTDG | 207 |
| DB | 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETVETKEITTTNKQSTDG | 300 |
| QY | 208 KBKTTSAKETQSIEKTSAXDLAFTSKVLAKPTPKAETTTKGPALTTPKGPTTTPKEPAS | 267 |
| DB | 301 KEKTTSAKETQSIEKTSAXDLAFTSKVLAKPTPKAETTTKGPALTTPKGPTTTPKEPAS | 360 |
| QY | 268 TTPKEPTTTIKSAPTTPKEPAPTTIKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP | 327 |
| DB | 361 TTPKEPTTTIKSAPTTPKEPAPTTIKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP | 420 |
| QY | 328 APITTSKAPTTKEPAPTTPKDAPTTPKAPAPTTTPKEPTPTTPKGPAPTTKEPAPTTPK | 387 |
| DB | 421 APITTSKAPTTKEPAPTTPKDAPTTPKAPAPTTTPKEPTPTTPKGPAPTTKEPAPTTPK | 480 |
| QY | 388 EPAPTAPKPAPTTPKEPAPTTTPKEPAPTTTPKSPSTTPKEPAPTTTPKSAPTTTKEPAPT | 447 |
| DB | 481 EPAPTAPKPAPTTPKEPAPTTTPKEPAPTTTPKSPSTTPKEPAPTTTPKSAPTTTKEPAPT | 540 |
| QY | 448 TTSAPTTPKEPPTTTKEPAPTTTPKEPAPTTTPKKAAPTTTPKEPAPTTTPKEPAPTTTPKP | 507 |
| DB | 541 TTSAPTTPKEPPTTTKEPAPTTTPKEPAPTTTPKKAAPTTTPKEPAPTTTPKEPAPTTTPKP | 600 |
| QY | 508 APAPKPAPTTPKETAPTTPKKLTPTTPPKLAPTTPKEPAPTTPEELAPTTPEEPTPTT | 567 |
| DB | 601 APAPKPAPTTPKETAPTTPKKLTPTTPPKLAPTTPKEPAPTTPEELAPTTPEEPTPTT | 660 |
| QY | 568 PEPAPPTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP | 627 |
| DB | 661 PEPAPPTPKAAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEP | 720 |
| QY | 628 APPTPKAPKELAPTTTKBPTSTSDKAPTTTPKGTAPTTPKEPAPTTTPKEPAPTTPKG | 687 |
| DB | 721 APPTPKAPKELAPTTTKBPTSTSDKAPTTTPKGTAPTTPKEPAPTTTPKEPAPTTPKG | 780 |
| QY | 688 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDKAPTTTPKETAPTTPKEPAPTTPK | 747 |
| DB | 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDKAPTTTPKETAPTTPKEPAPTTPK | 840 |
| QY | 748 KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSPPDESTPELSAETPKALENSKPEGVPT | 807 |
| DB | 841 KPAPTTPETPPPTTSVSTPTTTKEPTTIHKSPPDESTPELSAETPKALENSKPEGVPT | 900 |
| QY | 808 TKTPAATKPEMTTAKDKXTTERDLRTTPETTTAAAPKMTATTEKTTESKITATTITQV | 867 |
| DB | 901 TKTPAATKPEMTTAKDKXTTERDLRTTPETTTAAAPKMTATTEKTTESKITATTITQV | 960 |
| QY | 868 TSTTTQDTPPKITTLKTTTLAPKVTITTKXITTTTEIMNKPEETAKPKDRATNSKATTPK | 927 |
| DB | 961 TSTTTQDTPPKITTLKTTTLAPKVTITTKXITTTTEIMNKPEETAKPKDRATNSKATTPK | 1020 |

| | | |
|-----------|---|------|
| QY | 928 POKPTKAPKPTSTKKPTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRN | 987 |
| DB | 1021 POKPTKAPKPTSTKKPTMPRVKPKTTPTRKMTSTMPELNPTSRIAEAMLQTTTRN | 1080 |
| QY | 988 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIINPMLS | 1047 |
| DB | 1081 QTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLRPVNQGIINPMLS | 1140 |
| QY | 1048 DETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARRITEVWGIFSPIDTFT | 1107 |
| DB | 1141 DETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSFSPSPARRITEVWGIFSPIDTFT | 1200 |
| QY | 1108 RNCNCGKTPFFPKDSQVWRFTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPESVY | 1167 |
| DB | 1201 RNCNCGKTPFFPKDSQVWRFTNDIKDAGYKPKIFKFGGLTGQIVAAALSTAKYKNWPESVY | 1260 |
| QY | 1168 FFKRGSGIQOYIYKQEPVQKCFGRPALNYPVYGE | 1202 |
| DB | 1261 FFKRGSGIQOYIYKQEPVQKCFGRPALNYPVYGE | 1295 |
| RESULT 11 | | |
| AAB29778 | | |
| ID | AAB29778 standard; protein; 902 AA. | |
| XX | | |
| AC | AAB29778; | |
| XX | | |
| DT | 28-FEB-2001 (first entry) | |
| XX | | |
| DE | Human MSF-derived tribonection. | |
| XX | | |
| KW | Human tribonection; MSF; megakaryocyte stimulating factor; | |
| KW | alternative splicing; joint boundary lubricant; O-linked oligosaccharide; | |
| KW | osteoarthritis; tribosupplementation; tissue adhesion inhibition; | |
| KW | friction coefficient reduction; gene therapy; antiarthritic; osteopathic. | |
| OS | Homo sapiens. | |
| XX | | |
| PN | WO200064930-A2. | |
| XX | | |
| PD | 02-NOV-2000. | |
| XX | | |
| PF | 24-APR-2000; 2000WO-US010953. | |
| XX | | |
| PR | 23-APR-1999; 99US-00298970. | |
| XX | | |
| PA | (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER. | |
| XX | | |
| PI | Jay GD; | |
| XX | | |
| DR | WPI; 2001-024673/03. | |
| XX | | |
| PT | Novel tribonection polypeptide useful as lubricant for treating | |
| XX | osteoarthritis, comprises O-linked lubricating moiety. | |
| PS | Disclosure; Fig 1; 47pp; English. | |
| XX | | |
| CC | The invention relates to a human tribonection which is a product of | |
| CC | alternative splicing of the human MSF (megakaryocyte stimulating factor) | |
| CC | gene. The tribonection has at least one O-linked oligosaccharide | |
| CC | lubricating moiety and has a polypeptide sequence comprising 1-76 repeats | |
| CC | of a motif having at least 50% identity to the sequence KEPAPT | |
| CC | (AAB29774). The invention also relates to a nucleic acid encoding a human | |
| CC | MSF-derived tribonection; a biocompatible composition comprising a human | |
| CC | tribonection for inhibiting tissue adhesion formation; and a method of | |
| CC | diagnosing osteoarthritis or a predisposition to osteoarthritis by | |
| CC | measuring the amount of MSF or its fragment in a biological sample of a | |
| CC | mammal, wherein an increased amount of MSF compared to a control | |
| CC | indicates the presence of or predisposition to developing osteoarthritis. | |
| CC | The tribonection and DNA encoding it are useful in the treatment of | |
| CC | osteoarthritis, where they may be used for lubricating mammalian joints, | |
| CC | such as articulating joints of humans, dogs or horses. The tribonection, | |

CC when formulated as a membrane, foam, gel or fibre, is useful for
 CC inhibiting adhesion between two surfaces such as the injured tissues of a
 CC mammal, where the injury is caused by a surgical insertion or trauma, or
 CC an artificial device e.g., an orthopaedic implant. In particular, one of
 CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be
 CC used in gene therapy. The present sequence represents a substantial
 CC portion of a human MSP-derived tribonectin
 XX
 SQ Sequence 902 AA;

Query Match 50.3%; Score 3552; DB 4; Length 902;
 Best Local Similarity 74.4%; Pred. No. 3.2e-82;
 Matches 795; Conservative 27; Mismatches 76; Indels 170; Gaps 66;

QY 107 VKDNKNRKTGKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTTOHKNVSTSPKITTAKPIN 166
 DB 1 VKDNKNRKTGKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTTOHKNVSTSPKITTAKPIN 60

QY 167 PRSLPNSDTSKETSILTVNKETTVETKETTNTKQSTDGKEKTSKETSQSIKTSK 226
 DB 61 PRSLPNSDTSKETSILTVNKETTVETKETTNTKQSTDGKEKTSKETSQSIKTSK 120

QY 227 DLAPTSKVLAKPDKAETTTKGPALTTPKPEPTTTKPEASTTPKPTTTIKSAPTTPK 286
 DB 121 DLAPTSKVLAKPDKAETTTKGPALTTPKPEPTTTKPEASTTPKPTTTIKSAPTTPK 172

QY 287 EPAPTTKSAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTT 346
 DB 173 EPAPTTKSAPTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTT 228

QY 347 PKKPAPTTKEPAPTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAP 406
 DB 229 -KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT 281

QY 407 PTTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 466
 DB 282 PTTTKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTTPKE 335

QY 467 PAPTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTT 526
 DB 336 PAPTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTT 388

QY 527 TPKKLTTPPTPEKLAFTTPKAPTTTPEELAPTTTPEEPTTTPPEEAPTTTPKAAAPNTPK 586
 DB 389 T-KEPAPTTKEP-APTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KE 440

QY 587 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 646
 DB 441 PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 488

QY 647 EPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTPKEPAPTTTPKGPAP 706
 DB 489 EPAPTTPKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAP 541

QY 707 --KELAPTTTSGKPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTT 764
 DB 542 TTKEPAP-TTKEPAPTTKE-PAPTT-KEPAPTT-KEPAPTT-KEPAPTT-KEPAPTTKE-PAPTTPKE 595

QY 765 STPTTTTKEPTTHKSPDSELSREPTPKALENSP--KEPGVPTTKTPAATKPEMTTAA 822
 DB 596 A--PTTKEPAPTTKEPAPTT--KEPAPTTKEPAPTTTTPKEPAPTT--APTTPKEPAPTT----- 640

QY 823 KDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQVSTTTTODTTPFKLIT 882
 DB 641 -----TKEPAPTT-----KEPAPTTKEPAPTTTTPKEPAPTT----- 667

QY 883 LKTTTLAPKVTTTKTIITTTIMNKPETAKPKDRATNSKATTPKPKQPTKAPKKPTSTK 942
 DB 668 -----TTKEPAPTTTTPKEPAPTT-----TTKEPAPTTTTPKEPAPTT----- 684

QY 943 KPKTMPRVKPKTTTPPKMTSTMPELNPTSGRIAEAMLQTTTRNQTPNSKLVENPKSE 1002
 DB 685 EP-----APTTPTPKMTSTMPELNPTSGRIAEAML--TTTRNQTPNSKLVENPKSE 735

QY 1003 DAGAAGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLSDETNICNGKFPVDGLT 1062
 DB 736 DAGAAGETPHMLLRPHVFMPEVTPDMXYLPRVFNQGIINPMLSDETNICNGKFPVDGLT 795

QY 1063 TLRNGTLVAFRGHYFWMLSFSPSPARRITEVWGIPSPIDTPTTRCNCEGKTPFFKDSQ 1122
 DB 796 TLRNGTLVAFRGHYFWMLSFSPSPARRITEVWGIPSPIDTPTTRCNCEGKTPFFKDSQ 855

QY 1123 YWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFK 1170
 DB 856 YWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFK 902

RESULT 12

ABU53253

ID ABU53253 standard; protein; 551 AA.

XX AC ABU53253;

XX DT 14-APR-2003 (first entry)

XX DE Human testes-derived DKFZphtes3_4019 homologue #2.

XX KW Human; gene therapy; vaccine; disease treatment; detection.

XX OS Homo sapiens.

XX PN WO200112659-A2.

XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000WO-IB001496.

XX PR 18-AUG-1999; 99US-0149499P.

XX PR 28-SEP-1999; 99US-0156503P.

XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX PI Wiemann S;

XX DR WPI; 2001-327840/34.

XX PT Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.

XX PS Example III; Page 892-893; 1095pp; English.

XX CC This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

SQ Sequence 551 AA;

Query Match

Best Local Similarity 41.5%; Score 2929; DB 4; Length 551;

Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 504 TKKPAPTKAPKEPAPTTPKETAPTTPKLTTPPTPEKLAFTTPKEPAPTTPEELAPTTPEEP 563
 DB 1 TKKPAPTKAPKEPAPTTPKETAPTTPKLTTPPTPEKLAFTTPKEPAPTTPEELAPTTPEEP 60

QY 564 TPTTPEEPAPTTPKAAAPNTPKAPPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPT 623
 DB 61 TPTTPEEPAPTTPKAAAPNTPKAPPTTTPKEPAPTTTPKEPAPTTTPKGTAPT 120

QY 624 LKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKKEPAPTTKKEPAPT 683
DB 121 LKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKKEPAPTTKKEPAPT 180
QY 684 TPKGTAPTTLKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKKEPAP 743
DB 181 TPKGTAPTTLKEPAPTTKKPAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKKEPAP 240
QY 744 TTPKKPAPTTPTPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKEP 803
DB 241 TTPKKPAPTTPTPTPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKEP 300
QY 804 GVTPTTKTAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETATTTTEKTESKITAT 863
DB 301 GVTPTTKTAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETATTTTEKTESKITAT 360
QY 864 TTQVSTTTQDTPPKITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKA 923
DB 361 TTQVSTTTQDTPPKITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKA 420
QY 924 TTPKQKPTKAPKKPTSTKKPKTMPRVKPKTTPTRKMTSTMPELNPTSRJAEAMLQTT 983
DB 421 TTPKQKPTKAPKKPTSTKKPKTMPRVKPKTTPTRKMTSTMPELNPTSRJAEAMLQTT 480
QY 984 TRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIIN 1043
DB 481 TRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIIN 540
QY 1044 PMLSDETNI:CN 1054
DB 541 PMLSDETNI:CN 551
RESULT 13
ID ABUS3252 standard; protein; 546 AA.
XX AC ABUS3252;
DT XX
XT 14-APR-2003 (first entry)
DE Human testes-derived DKFZphtes3_4o19 homologue #1.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX FT Nucleic acids having the sequences of clones isolated from libraries of
XX PT different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 892; 1095pp; English.
XX CC This invention describes novel polynucleotides and polypeptides isolated
XX CC from human cDNA libraries which can be used for gene therapy or in
XX CC vaccines. The polynucleotides of the invention and antibodies encoded by
XX CC them may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate polypeptide expression. The products of the
XX CC invention may also be used to identify modulators of expression and

CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 546 AA;
Query Match 41.3%; Score 2920; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 1.8e-66;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 395 KKPAPTTTKKEPAPTTKKEPAPTTTKEPSPPTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 454
DB 1 KKPAPTTKEPAPTTKKEPAPTTTKEPSPPTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 60
QY 455 TPKEPSPPTTKEPAPTTKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 514
DB 61 TPKEPSPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 120
QY 515 PAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTTPEEPAPT 574
DB 121 PAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEEPTPTTPEEPAPT 180
QY 575 TPKAAAPNTPKPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKGTAPTTLKEPAPTTPKK 634
DB 181 TPKAAAPNTPKPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKGTAPTTLKEPAPTTPKK 240
QY 635 PAKKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKKEPAPTTPKKEPAPTTPKGTAPTTLK 694
DB 241 PAKKELAPTTTKEPTSTSDKAPATTTPKGTAPTTKKEPAPTTPKKEPAPTTPKGTAPTTLK 300
QY 695 EPAPTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKETAPTTPKKEPAPTTPKKAPTTTP 754
DB 301 EPAPTTPKKAPKELAPTTTKEPTSTSDKAPATTTPKETAPTTPKKEPAPTTPKKAPTTTP 360
QY 755 ETTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKEGVPTTKPAAT 814
DB 361 ETTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAETPKALENSPKEGVPTTKPAAT 420
QY 815 KPEMTTAKDKTTERDLRTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTQD 874
DB 421 KPEMTTAKDKTTERDLRTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTQD 480
QY 875 TTPEKJITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPKQKPTKA 934
DB 481 TTPEKJITLTKTTLAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPKQKPTKA 540
QY 935 PKKPTS 940
DB 541 PKKPTS 546
RESULT 14
AAO18834
ID AAO18834 standard; protein; 538 AA.
XX AC AAO18834;
XX DT 29-OCT-2002 (first entry)
XX DE
XX DE
XX DE
XX KW 3' cartilage superficial zone protein coding sequence encoded protein.
XX KW SZP; superficial zone protein; cartilage; lubrication; human;
XX KW degenerative joint condition; arthritis; osteoporosis; trauma; CACF;
XX KW chondroicin sulphate substitution consensus; antiarthritic;
XX KW antirheumatic; osteopathic; antigout; antiinflammatory; dermatological;
XX KW immunosuppressive.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 114 /note= "encoded by ACTACT"

Wed Oct 13 12:38:28 2004

```

|||
241 TPKEPAPTTKEPTTTPKEPAPTTKEPAPTTAPKCPAPTTKEPAPTTTPKEP 300
QY
414 APTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTTSAPTTKEPSPTTTPKEPAPTTTPK 473
Db
301 APTTTKEPSPTTPKEPAPTTTKSAPTTTKEPAPTTTTSAPTTKEPSPTTTPKEPAPTTTPK 360
QY
474 EPAPTTPKKCPAPTTTPKEPAPTTTPKEPAPTTTTPKCPAPTTAPKEPAPTTTPKETAPTTPKKLTP 533
Db
361 EPAPTTPKKCPAPTTTPKEPAPTTTPKEPAPTTTTPKCPAPTTAPKEPAPTTTPKETAPTTPKKLTP 420
QY
534 TTPEKLAPTTPKCPAPTTTPBEELAPTTTPBEPTTTPBEPTTTPBEPTTTPBEPTTTPBEPTTTPK 593
Db
421 TTPEKLAPTTPKCPAPTTTPBEELAPTTTPBEPTTTPBEPTTTPBEPTTTPBEPTTTPBEPTTTPK 480
QY
594 EPAPTTKEPAPTTTPKETAPTTPKGTAPTTLKE 626
Db
481 EPAPTTKEPAPTTTPKETAPTTPKGTAPTTLKE 513
```

Search completed: October 13, 2004, 11:37:31
Job time : 114.696 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 22.7891 Seconds
(without alignments)
3815.116 Million cell updates/sec

Title: SEQ1-H

Perfect score: 7064

Sequence: 1 MAWKTLPIVILLLSVFVIQ.....ARAITRSGQTLSKWNVNC 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents:AA:*

- 1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/PCJUS COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 7064 | 100.0 | 1311 | 4 | US-07-757-022B-42 |
| 2 | 7049.7 | 99.8 | 1354 | 4 | US-07-757-022B-48 |
| 3 | 7049 | 99.8 | 1361 | 4 | US-07-757-022B-40 |
| 4 | 7044.7 | 99.7 | 1404 | 4 | US-07-757-022B-2 |
| 5 | 7044.7 | 99.7 | 1404 | 4 | US-07-757-022B-62 |
| 6 | 7044.7 | 99.7 | 1404 | 4 | US-09-298-970A-1 |
| 7 | 7029.7 | 99.5 | 1404 | 4 | US-10-164-595-78 |
| 8 | 6811.3 | 96.4 | 1314 | 4 | US-07-757-022B-50 |
| 9 | 6799.9 | 96.3 | 1270 | 4 | US-07-757-022B-44 |
| 10 | 6785.6 | 96.1 | 1313 | 4 | US-07-757-022B-142 |
| 11 | 6784.9 | 96.0 | 1320 | 4 | US-07-757-022B-46 |
| 12 | 6784.9 | 96.0 | 1320 | 4 | US-07-757-022B-60 |
| 13 | 6780.6 | 96.0 | 1363 | 4 | US-07-757-022B-52 |
| 14 | 6769.9 | 95.8 | 1320 | 4 | US-10-164-595-58 |
| 15 | 5603.7 | 79.3 | 1140 | 4 | US-07-757-022B-104 |
| 16 | 5500.7 | 77.9 | 1038 | 4 | US-07-757-022B-74 |
| 17 | 5472.6 | 77.5 | 1022 | 4 | US-07-757-022B-84 |
| 18 | 5344.6 | 75.7 | 1049 | 4 | US-07-757-022B-58 |
| 19 | 5011 | 70.9 | 941 | 4 | US-07-757-022B-14 |
| 20 | 2057.8 | 29.1 | 463 | 4 | US-07-757-022B-54 |
| 21 | 1806.8 | 25.6 | 423 | 4 | US-07-757-022B-66 |
| 22 | 1793.7 | 25.4 | 422 | 4 | US-07-757-022B-68 |
| 23 | 1723.1 | 24.4 | 372 | 4 | US-07-757-022B-64 |
| 24 | 1398.2 | 19.8 | 5179 | 4 | US-09-538-092-1258 |
| 25 | 1219.6 | 17.3 | 237 | 4 | US-07-757-022B-72 |
| 26 | 1139.7 | 16.1 | 8991 | 4 | US-08-714-741-32 |
| 27 | 1108.5 | 15.7 | 296 | 4 | US-07-757-022B-70 |

28 884 12.5 157 4 US-07-757-022B-102
 29 884 12.5 157 4 US-07-757-022B-114
 30 869 12.3 207 4 US-07-757-022B-116
 31 869 12.3 207 4 US-07-757-022B-136
 32 849.6 12.0 3118 3 US-09-579-181-1
 33 842.3 11.9 2972 3 US-09-579-181-2
 34 824.7 11.7 3256 4 US-09-919-172-98
 35 824.7 11.7 3256 4 US-09-976-594-22
 36 814.8 11.5 1837 3 US-09-919-039-21
 37 814.8 11.5 1837 3 US-08-928-361B-5
 38 814.8 11.5 1837 4 US-09-588-995A-5
 39 808.9 11.5 2142 4 US-09-538-092-1142
 40 808.4 11.4 4019 4 US-09-854-133-425
 41 794.6 11.2 1721 3 US-08-700-651-5
 42 794.6 11.2 1721 3 US-08-928-361B-6
 43 794.6 11.2 1721 4 US-09-588-995A-6
 44 771 10.9 231 4 US-07-757-022B-30
 45 763 10.8 132 4 US-07-757-022B-140

ALIGNMENTS

RESULT 1
 US-07-757-022B-42
 ; Sequence 42, Application US/07757022B
 ; Patent No. 6433142
 ; GENERAL INFORMATION:
 ; APPLICANT: Gesner, Thomas G.
 ; APPLICANT: Clark, Stephen C.
 ; APPLICANT: Turner, Katherine
 ; APPLICANT: Hewick, Rodney M.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07757,022B
 ; FILING DATE: 19910910
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)876-1170
 ; TELEFAX: (617)876-5851
 ; INFORMATION FOR SEQ ID NO: 42:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1311 amino acids
 ; TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-42

Query Match 100.0%; Score 7064; DB 4; Length 1311;
Best Local Similarity 100.0%; Pred. No. 4.5e-193; Indels 0; Gaps 0;
Matches 1311; Conservative 0; Mismatches 0;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSRDATCNCYNQHYMECCPDF 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSRDATCNCYNQHYMECCPDF 60

QY 61 KRVCVTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVKDNKKQVKKKPT 120
DB 61 KRVCVTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYSEFCAEVKDNKKQVKKKPT 120

QY 121 PKPVPVDEAGSLGNDGKVTTPDSTTQHNKVSTSPKLTAKPINRPSLPNSDTSKE 180
DB 121 PKPVPVDEAGSLGNDGKVTTPDSTTQHNKVSTSPKLTAKPINRPSLPNSDTSKE 180

QY 181 TSLTVNKEITTVETKETITTNKQSTDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTP 240
DB 181 TSLTVNKEITTVETKETITTNKQSTDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTP 240

QY 241 KAETTTKGALATTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTP 300
DB 241 KAETTTKGALATTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTP 300

QY 301 KEPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTP 360
DB 301 KEPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTP 360

QY 361 TTPKEPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTP 420
DB 361 TTPKEPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTP 420

QY 421 PSPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 480
DB 421 PSPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 480

QY 481 KKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 540
DB 481 KKPAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 540

QY 541 PTTPEKPAPTTPPEELAPTTPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPE 600
DB 541 PTTPEKPAPTTPPEELAPTTPEEPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTPPE 600

QY 601 KEPAPTTPKETAPTTPKGATPTTLKEPAPTTPKKPAKELAPTTKEPTSTTSKDPAPT 660
DB 601 KEPAPTTPKETAPTTPKGATPTTLKEPAPTTPKKPAKELAPTTKEPTSTTSKDPAPT 660

QY 661 PKGTAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 720
DB 661 PKGTAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 720

QY 721 TSDKPAPTTPKETAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 780
DB 721 TSDKPAPTTPKETAPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 780

QY 781 DESTPELSAETTPKALSNSKEPGVPTTKPAATKPEMTTITAKDKTTERDLRTTPETTTA 840
DB 781 DESTPELSAETTPKALSNSKEPGVPTTKPAATKPEMTTITAKDKTTERDLRTTPETTTA 840

QY 841 APKMTKETATTTETKTESKITATTTQVSTSTTQDTPFKITLTKTTLAPKVTTKKIT 900
DB 841 APKMTKETATTTETKTESKITATTTQVSTSTTQDTPFKITLTKTTLAPKVTTKKIT 900

QY 901 TTEIMNKPEETAKPKORATNSKATTPPKQKTKAPKPTSTKKPKTPRVRKPKTTPPR 960
DB 901 TTEIMNKPEETAKPKORATNSKATTPPKQKTKAPKPTSTKKPKTPRVRKPKTTPPR 960

QY 961 KWTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENPKSEDAGGABGETPHMLLRPHV 1020

RESULT 2

US-07-757-022B-48
; Sequence 48, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids

```
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-48

Query Match          99.8%; Score 7049.7; DB 4; Length 1354;
Best Local Similarity 96.8%; Pred. No. 1.2e-192;
Matches 1311; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSLVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLSLVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCOHYMECCPDF 60
QY 61 KRVCYTAELSCGRCFCFSFERGRCDCDAQCKYDKCCPDYSEFCAE-----106
DB 61 KRVCYTAELSCGRCFCFSFERGRCDCDAQCKYDKCCPDYSEFCAEHSYSEVQESSSS 120
QY 107 -----VKONKQRTKKKPTKPPVVDGAGSLDNGD 137
DB 121 SSSSSSTIWKIKSSKNSAANRELQKLVKONKQRTKKKPTKPPVVDGAGSLDNGD 180
QY 138 FKVTTPDSTTOHNVKYSTPKITAKPINRPDLPPNSDTSKETSLSLVNKEITVETKETT 197
DB 181 FKVTTPDSTTOHNVKYSTPKITAKPINRPDLPPNSDTSKETSLSLVNKEITVETKETT 240
QY 198 TTNKQTSQDGEKNTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPAITTPKEP 257
DB 241 TTNKQTSQDGEKNTSAKETQSIEKTSKDLAPTSKVLAKPTPKAETTTKGPAITTPKEP 300
QY 258 TPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTTKBPAPTTTKBPAPTTPK 317
DB 301 TPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTTKBPAPTTTKBPAPTTPK 360
QY 318 EPAPTTTKBPAPTTTKSAPTTTPKEPAPTTPKBPAPTTTPKEPAPTTTPKEPAPTT 377
DB 361 EPAPTTTKBPAPTTTKSAPTTTPKEPAPTTPKBPAPTTTPKEPAPTTTPKEPAPTT 420
QY 378 TKEPAPTTTPKEPAPTTAKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSA 437
DB 421 TKEPAPTTTPKEPAPTTAKKPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSA 480
QY 438 PTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 497
DB 481 PTTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTPK 540
QY 498 EPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKLPTTPKEPAPTTTPKEPAPTTPEELAP 557
DB 541 EPAPTTTKKPAPTAPKEPAPTTPKETAPTTPKLPTTPKEPAPTTTPKEPAPTTPEELAP 600
QY 558 TTPPEPTTPTPKEPAPTTPKAAAPNTPKBPAPTTTPKEPAPTTTPKEPAPTTPKETAPTTPK 617
DB 601 TTPPEPTTPTPKEPAPTTPKAAAPNTPKBPAPTTTPKEPAPTTTPKEPAPTTPKETAPTTPK 660
QY 618 GTAPTTTPKEPAPTTTPKPAKELAPTTTPKEPAPTTSDKPAPTTPKGTAPTTTPKEPAPTT 677
DB 661 GTAPTTTPKEPAPTTTPKPAKELAPTTTPKEPAPTTSDKPAPTTPKGTAPTTTPKEPAPTT 720
QY 678 KEPAPTTTPKGTAPTTTPKPAKELAPTTTPKGTSTTSKDPAPTTTPKEPAPTT 737
DB 721 KEPAPTTTPKGTAPTTTPKPAKELAPTTTPKGTSTTSKDPAPTTTPKEPAPTT 780
QY 738 PKEPAPTTTPKPAPTTPPTTPPTTSEVSTPTTPKEPTTIHKSPDSESTPELSAETPKALE 797
DB 781 PKEPAPTTTPKPAPTTPPTTPPTTSEVSTPTTPKEPTTIHKSPDSESTPELSAETPKALE 840
QY 798 NSPKPGVPTTKPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKEATTTTEKTE 857
DB 841 NSPKPGVPTTKPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKEATTTTEKTE 900
QY 858 SKITATTTQVTSSTTQDTPPKITTLKTTTLAPKVTTKTKTITTEIMNKPEETAKPKDR 917
DB 901 SKITATTTQVTSSTTQDTPPKITTLKTTTLAPKVTTKTKTITTEIMNKPEETAKPKDR 960

RESULT 3
US-07-757-022B-40
; Sequence 40, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
```



```

; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-2

Query Match          99.7%; Score 7044.7; DB 4; Length 1404;
Rest Local Similarity 93.4%; Pred. No. 1.8e-192;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIVILLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYWECPCDF 60
DB 1 MAWKTLPIVILLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCYDNCQHYWECPCDF 60
QY 61 KRVCIAELSCGRCPSPRGRECDCAQCKYDKCCPDYESFCAE----- 106
DB 61 KRVCIAELSCGRCPSPRGRECDCAQCKYDKCCPDYESFCAE----- 106
QY 107 ----- 106
DB 121 PFGASQTIKSTTKSRPKPKPNKKTKVIESEITEHSVSENQESSSSSSSSSTI 180
QY 107 -----VKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTPTDTST 147
DB 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTPTDTST 240
QY 148 TQHNKVSIPKLTAKINPRSLPNSDTSKETSITVNKETTVEKETTITNKQISTDG 207
DB 241 TQHNKVSIPKLTAKINPRSLPNSDTSKETSITVNKETTVEKETTITNKQISTDG 300
QY 208 KEKTTSAKETQSIKTSKDLAPTQKVLAKTPKAEITTKGALITPKPEPTTTKPEKAS 267
DB 301 KEKTTSAKETQSIKTSKDLAPTQKVLAKTPKAEITTKGALITPKPEPTTTKPEKAS 360
QY 268 TPKPEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 327
DB 361 TPKPEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 420
QY 328 APITTKSAPTTKPEAPTTPKKPAPTTPKPEAPTTKPEPTTTTPKPEAPTTKPEAPTTPK 387
DB 421 APITTKSAPTTKPEAPTTPKKPAPTTPKPEAPTTKPEPTTTTPKPEAPTTKPEAPTTPK 480
QY 388 EPAPTAPKPAPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 447
DB 481 EPAPTAPKPAPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 540
QY 448 TTKSAPTTKPEPSTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 507
DB 541 TTKSAPTTKPEPSTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKKP 600
QY 508 APTAPKPEAPTTKPEAPTTTPKLTTPTEKLAPTTPKPEAPTTPEELAPTTPEEPTPTT 567
DB 601 APTAPKPEAPTTKPEAPTTTPKLTTPTEKLAPTTPKPEAPTTPEELAPTTPEEPTPTT 660
QY 568 PEEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 627
DB 661 PEEAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 720
QY 628 APITPKKPAKELAPTTTKETSTSDKPAPTTPKGTAPTTKPEAPTTKPEAPTTKPE 687
DB 721 APITPKKPAKELAPTTTKETSTSDKPAPTTPKGTAPTTKPEAPTTKPEAPTTKPE 780
QY 688 TAPITLKPEAPTTPKKPAKELAPTTTKGPTSTSDKPAPTTPKETAPTTKPEAPTTPK 747

```

RESULT 5

```

US-07-757-022B-62
; Sequence 62, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: AMINO ACIDS
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-62

Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
Best Local Similarity 93.4%; Pred. No. 1.8e-192;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIVLLLLSVFVIQVSSODLSSCAGRCGEGYSDATCNCYNCQHYMECCPDF 60
QY 61 KRVCVTAELCKGRCFESFERGECDCDAQCKYDKCCPDYEFSCAE----- 106
DB 61 KRVCVTAELCKGRCFESFERGECDCDAQCKYDKCCPDYEFSCAEVHNFTSPSSKKAP 120
QY 107 ----- 106
DB 121 PPSGASQTKSTTKRSPKPPNKKTKVLEBEITBEHSVSENQESSSSSSSSSTIW 180
QY 107 -----VKONKKNRKTKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 147
DB 181 KIKSSKNSAANRELQKKVKONKKNRKTKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 148 TQHKVSTGKITTAKPINRPSLPNSDTSKETSLSLVNKEITVEKETTINKQSTDG 207
DB 241 TQHKVSTGPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVEKETTINKQSTDG 300
QY 208 KEKTTSAKETOSTEKTSADKLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTPPKEPAS 267
DB 301 KEKTTSAKETOSLEKTSADKLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTPPKEPAS 360
QY 268 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTKPEPAPTTTKPEP 327
DB 361 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEP 420
QY 328 APITTSKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTK 387
DB 421 APITTSKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTK 480
QY 388 EPAPTAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKSAPTTTKPEPAPT 447
DB 481 EPAPTAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKSAPTTTKPEPAPT 540
QY 448 TTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKKXP 507
DB 541 TTKSAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTTKKXP 600
QY 508 APITAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTPEELAPTTPEEPPTT 567
DB 601 APITAPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTPEELAPTTPEEPPTT 660
QY 568 PEEPAPTTPKAAAPNTKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKGTAPTTLKEP 627

DB 661 PEEPAPTTPKAAAPNTKPEPAPTTPKPEPAPTTPKPEPAPTTPKGTAPTTLKEP 720
QY 628 APITPKKPAKELAPTTTKETSTTSDDKPAAPTTPKGTAPTTPKBPAPTTPKBPAPTTPKG 687
DB 721 APITPKKPAKELAPTTTKETSTTSDDKPAAPTTPKGTAPTTPKBPAPTTPKBPAPTTPKG 780
QY 688 TAPTTLKEPAPTTPKKPAKELAPTTTKGPTSTTSDDKPAAPTTPKGTAPTTPKBPAPTTPK 747
DB 781 TAPTTLKEPAPTTPKKPAKELAPTTTKGPTSTTSDDKPAAPTTPKGTAPTTPKBPAPTTPK 840
QY 748 KPAAPTTPEPTTSEVSTPTTTKEPTTIHKSPDSESTBELSAEPTPKALENSPKPGVPT 807
DB 841 KPAAPTTPEPTTSEVSTPTTTKEPTTIHKSPDSESTBELSAEPTPKALENSPKPGVPT 900
QY 808 TKTPAATKPEMTTAKDKTTEDLATTPTETTTAAKPKMKETATTTKTTESKITATTITQV 867
DB 901 TKTPAATKPEMTTAKDKTTEDLATTPTETTTAAKPKMKETATTTKTTESKITATTITQV 960
QY 868 TSTTTQDTPPKITTLTKTTLLAPKVTTKKIIITTEIMNKPETAKPKDRATNSKATTPK 927
DB 961 TSTTTQDTPPKITTLTKTTLLAPKVTTKKIIITTEIMNKPETAKPKDRATNSKATTPK 1020
QY 928 POKPTKAPKKPTSTKKPKTMPRVKPKTPTPRKMTSTMPELNPTSRIAEAMLOTTTRN 987
DB 1021 POKPTKAPKKPTSTKKPKTMPRVKPKTPTPRKMTSTMPELNPTSRIAEAMLOTTTRN 1080
QY 988 QTPNSKLEVEVNPKSDAGAGETPHMLLRPHVMEVETPDMDYLPRVNPQGIINPMLS 1047
DB 1081 QTPNSKLEVEVNPKSDAGAGETPHMLLRPHVMEVETPDMDYLPRVNPQGIINPMLS 1140
QY 1048 DETNICNGKVPDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPSPIDTFT 1107
DB 1141 DETNICNGKVPDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPSPIDTFT 1200
QY 1108 RNCSEKTPFFKDSQWRFRTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY 1167
DB 1201 RNCSEKTPFFKDSQWRFRTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY 1260
QY 1168 EFKRGGSIQOYLYKQSPVOKCGRRPALNVPVVGEMTQVRRRRFERAIGPSQHTTIRIQY 1227
DB 1261 EFKRGGSIQOYLYKQSPVOKCGRRPALNVPVVGEMTQVRRRRFERAIGPSQHTTIRIQY 1320
QY 1228 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1287
DB 1321 SPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYNNIDV 1380
QY 1288 PSTARAITTRSGQTLISKVWYNCP 1311
DB 1381 PSTARAITTRSGQTLISKVWYNCP 1404

RESULT 6
US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: Jav, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-298-970A-1

Query Match 99.7%; Score 7044.7; DB 4; Length 1404;
Best Local Similarity 93.4%; Pred. No. 1.8e-192;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAE----- 106
Db 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAE VHNTPSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHVSSENOESSSSSSSSSTIR 180
QY 107 -----VKDNKNRNTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTST 147
Db 181 KIKSSKNSAANRELQKKLVKDNKNRNTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTST 240
QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTSDG 207
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTSDG 300
QY 208 KEKTTSAKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTPKEPAS 267
Db 301 KEKTTSAKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTPKEPAS 360
QY 268 TTPKEPTTTPKSAPTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 327
Db 361 TTPKEPTTTPKSAPTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 420
QY 328 APPTTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 387
Db 421 APPTTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 480
QY 388 EPAPTAPKPPAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 447
Db 481 EPAPTAPKPPAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 540
QY 448 TTKSAPTTTPKSPPTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 507
Db 541 TTKSAPTTTPKSPPTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 600
QY 508 APPTAPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 567
Db 601 APPTAPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPK 660
QY 568 PEEAPTTTPKAAAPNTPKAPKELAPTTTPKGTSTTSKAPTTTPKEPATTTTPKEPATTTTPK 627
Db 661 PEEAPTTTPKAAAPNTPKAPKELAPTTTPKGTSTTSKAPTTTPKEPATTTTPKEPATTTTPK 720
QY 628 APPTPKAPKELAPTTTPKEPATTTTPKGTSTTSKAPTTTPKEPATTTTPKEPATTTTPK 687
Db 721 APPTPKAPKELAPTTTPKEPATTTTPKGTSTTSKAPTTTPKEPATTTTPKEPATTTTPK 780
QY 688 TAPTTTPKEPATTTTPKAPKELAPTTTPKGTSTTSKAPTTTPKEPATTTTPKEPATTTTPK 747
Db 781 TAPTTTPKEPATTTTPKAPKELAPTTTPKGTSTTSKAPTTTPKEPATTTTPKEPATTTTPK 840
QY 748 KPAPTTTPETPPPTSEVSTPTTKETPTTIHKSPPDESTPELSAETPKALENSPKERGVP 807
Db 841 KPAPTTTPETPPPTSEVSTPTTKETPTTIHKSPPDESTPELSAETPKALENSPKERGVP 900
QY 808 TKTPAATKPEMTTAKDTERDLRTPTTTPAATPKMTKETATTTTEKTTESKITATTTQV 867
Db 901 TKTPAATKPEMTTAKDTERDLRTPTTTPAATPKMTKETATTTTEKTTESKITATTTQV 960
QY 868 TSTTTQDTPPKITTLTKTTILAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPK 927
Db 961 TSTTTQDTPPKITTLTKTTILAPKVTTTKITITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 928 PQKPTAPKPKSTSKPKTWPVRVKPTTTPRKMSTMPELNPTSRIAEAMLTITTRPN 987
Db 1021 PQKPTAPKPKSTSKPKTWPVRVKPTTTPRKMSTMPELNPTSRIAEAMLTITTRPN 1080

QY 988 QTPNSKLVEVNPXSSEDAGAGETPHMLLRPHVPMPEVTPDMDYLPVPNQGIINPMLS 1047
Db 1081 QTPNSKLVEVNPXSSEDAGAGETPHMLLRPHVPMPEVTPDMDYLPVPNQGIINPMLS 1140
QY 1048 DETNINCNGKPVDCGLTTLRNGTLVAFRGHYFWMILSPSPSPARRITEVWIGIDPDIPTVFT 1107
Db 1141 DETNINCNGKPVDCGLTTLRNGTLVAFRGHYFWMILSPSPSPARRITEVWIGIDPDIPTVFT 1200
QY 1108 RCNCEGKTTFFKDSQYWRFTNDIKDAGYKPKPIKFGGLTGQIVAAALSTAKYKNWPESVY 1167
Db 1201 RCNCEGKTTFFKDSQYWRFTNDIKDAGYKPKPIKFGGLTGQIVAAALSTAKYKNWPESVY 1260
QY 1168 FFKRGSSIOQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFPERAIGESQTHHTIRIQY 1227
Db 1261 FFKRGSSIOQYIYKQEPVQKCPGRRPALNYPVYGMTQVRRRRFPERAIGESQTHHTIRIQY 1320
QY 1228 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYINIDV 1287
Db 1321 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYINIDV 1380
QY 1288 PSRTARAITRRSGQTLISKVYNCP 1311
Db 1381 PSRTARAITRRSGQTLISKVYNCP 1404

RESULT 7

US-10-164-595-78
; Sequence 78, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-78

Query Match 99.5%; Score 7029.7; DB 4; Length 1404;
Best Local Similarity 93.2%; Pred. No. 4.7e-192;
Matches 1308; Conservative 0; Mismatches 3; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAE----- 106
Db 61 KRVTCTAELCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAE VHNTPSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHVSSENOESSSSSSSSSTIR 180
QY 107 -----VKDNKNRNTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTST 147
Db 181 KIKSSKNSAANRELQKKLVKDNKNRNTKKKPTPKPPVVDGAGSLDNGDFKVTTPDTST 240
QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTSDG 207
Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTSDG 300
QY 208 KEKTTSAKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTPKEPAS 267
Db 301 KEKTTSAKETSQIEKTSADLAPTSKVLAKPTPKAETTTKGBALTTTPKEPTTTPKEPAS 360
QY 268 TTPKEPTTTPKSAPTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTKP 327

Db 1220 ISLNIRKPDGVDYVAFSKDQYINIDVPSRTARAITTRSGQTLSKVWYNCP 1270

RESULT 10
US-07-757-022B-142
; Sequence 142, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-142

Query Match 96.1%; Score 6785.6; DB 4; Length 1313;
Best Local Similarity 93.8%; Pred. No. 3.7e-185;
Matches 1270; Conservative 0; Mismatches 84; Gaps 2;

Qy 1 MAWKTLPIYLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLSVFVIQQVSSQ----- 25
Qy 61 KRVCUTABLSCXGRCFESFERGREGCDCAQCKKYDKCCPDYSEFCAE----- 106
Db 26 -----ELSCXGRCFESFERGREGCDCAQCKKYDKCCPDYSEFCAEHSVSENESSSSS 79
Qy 107 -----VKDNKKNTTKKKTTPKPPVVDGAGSLDNGD 137
Db 80 SSSSSSTTWIKSKSKNSANRELQKLVKDNKKNTTKKKTTPKPPVVDGAGSLDNGD 139

Qy 181 TSLVNVKETTIVETKETTINKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKPTP 240
Db 140 TSLVNVKETTIVETKETTINKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKPTP 199
Qy 241 KAEITTKGPAITTPKEPTTPKEBPASTTPKEPTPTTIKSAPTTPKEPAPITTKSAPTTP 300
Db 200 KAEITTKGPAITTPKEPTTPKEBPASTTPKEPTPTTIKSAPTTPKEPAPITTKSAPTTP 259
Qy 301 KEPAPITTKKEPAPITTPKEPAPITTKSAPTTPKEPAPITTPKKEPAPITTPKEPAP 360
Db 260 KEPAPITTKKEPAPITTPKEPAPITTKSAPTTPKEPAPITTPKKEPAPITTPKEPAP 319
Qy 361 TTPKEPTTPTPKEPAPITTPKEPAPITTPKAPKAPAPITTPKEPAPITTPKEPAPITTK 420
Db 320 TTPKEPTTPTPKEPAPITTPKEPAPITTPKAPKAPAPITTPKEPAPITTPKEPAPITTK 379
Qy 421 PSPTTPKEPAPITTKSAPTTPKEPAPITTKSAPTTPKEPAPITTPKEPAPITTPKEPAP 480
Db 380 PSPTTPKEPAPITTKSAPTTPKEPAPITTKSAPTTPKEPAPITTPKEPAPITTPKEPAP 439
Qy 481 KKPAPITTPKEPAPITTPKEPAPITTKKAPAPAPITTPKEPAPITTPKKTTPTPPEKLA 540
Db 440 KKPAPITTPKEPAPITTPKEPAPITTKKAPAPAPITTPKEPAPITTPKKTTPTPPEKLA 499
Qy 541 PTPPEKAPITTPPEEAPITTPPEEAPITTPPEEAPITTPKAAAPNTTPKEPAPITTPKEPAP 600
Db 500 PTPPEKAPITTPPEEAPITTPPEEAPITTPPEEAPITTPKAAAPNTTPKEPAPITTPKEPAP 559
Qy 601 KEPAPITTPKETAPITTPKETAPITTPKAPKAPKELAPITTKPTSTTSKDPAPITTP 660
Db 560 KEPAPITTPKETAPITTPKETAPITTPKAPKAPKELAPITTKPTSTTSKDPAPITTP 619
Qy 661 PKGTAPITTPKEPAPITTPKEPAPITTPKGTAPITTPKAPKAPKELAPITTKPTST 720
Db 620 PKGTAPITTPKEPAPITTPKEPAPITTPKGTAPITTPKAPKAPKELAPITTKPTST 679
Qy 721 TSDKAPITTPKETAPITTPKETAPITTPKAPKAPITTPPEPPTTSVSPPTTKETTHKSP 780
Db 680 TSDKAPITTPKETAPITTPKETAPITTPKAPKAPITTPPEPPTTSVSPPTTKETTHKSP 739
Qy 781 DESTPELSAEPTPKALENSKBPVPTTKPAATKPEMTTTAKDKITERDLRTPETTTA 840
Db 740 DESTPELSAEPTPKALENSKBPVPTTKPAATKPEMTTTAKDKITERDLRTPETTTA 799
Qy 841 APRWTKETAITTEKTTESKITATTTQVTSITTTQDTTPFKITLTKTTLAPKVTITTKIT 900
Db 800 APRWTKETAITTEKTTESKITATTTQVTSITTTQDTTPFKITLTKTTLAPKVTITTKIT 859
Qy 901 TTEIMNKPEBETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKKPMRVRKPKTTPTPR 960
Db 860 TTEIMNKPEBETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKKPMRVRKPKTTPTPR 919
Qy 961 KMTSTMPELNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSDEAGGAEGETPHMLRPHV 1020
Db 920 KMTSTMPELNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSDEAGGAEGETPHMLRPHV 979
Qy 1021 FMEPEVTPDMYLPVNPQGHIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWML 1080
Db 980 FMEPEVTPDMYLPVNPQGHIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWML 1039
Qy 1081 SPSPSPSPARRITEVWGIIPSPIDVFTRCNCEGKTFEFKOSQYWRFTNDIKDAGYKPIF 1140
Db 1040 SPSPSPSPARRITEVWGIIPSPIDVFTRCNCEGKTFEFKOSQYWRFTNDIKDAGYKPIF 1099
Qy 1141 KGFGGLTGQIVAAALSTAKYKNWPSVFFFKRGSGSIQQYIYKQEPVQKCGRRPALNVPVY 1200
Db 1100 KGFGGLTGQIVAAALSTAKYKNWPSVFFFKRGSGSIQQYIYKQEPVQKCGRRPALNVPVY 1159
Qy 1201 GEMTQVRRRRPERAIGBSQTHIRIQYSFARLAYQDKGVLNHNVKVSILWRGLPNVVTSA 1260
Db 1160 GEMTQVRRRRPERAIGBSQTHIRIQYSFARLAYQDKGVLNHNVKVSILWRGLPNVVTSA 1219
Qy 1261 ISLNIRKPDGVDYVAFSKDQYINIDVPSRTARAITTRSGQTLSKVWYNCP 1311

QY 138 FKVTPDSTQHNVKSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEITTVETKETT 197
Db 140 FKVTPDSTQHNVKSTSPKITTAKPINRPSLPPNSDTSKETSLSLVNKEITTVETKETT 199
QY 198 TTNKQTSIDGKEKTSKETSQISIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTTPKEP 257
Db 200 TTNKQTSIDGKEKTSKETSQISIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTTPKEP 259
QY 258 TPPTPKPASTTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTPK 317
Db 260 TPPTPKPASTTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTPK 319
QY 318 EPAPTTTKPEAPTTTKSAPTTKPEAPTTTKKAPTTTPKPEAPTTTPKPEAPT 377
Db 320 EPAPTTTKPEAPTTTKSAPTTKPEAPTTTKKAPTTTPKPEAPTTTPKPEAPT 379
QY 378 TKPEAPTTKPEAPTTAPKAPTTKPEAPTTKPEAPTTTKBSPPTTPKPEAPTTTKSA 437
Db 380 TKPEAPTTKPEAPTTAPKAPTTKPEAPTTKPEAPTTTKBSPPTTPKPEAPTTTKSA 439
QY 438 PTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKAPTTTPKPEAPTTTPK 497
Db 440 PTTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKKAPTTTPKPEAPTTTPK 499
QY 498 EPAPTTTKKAPTTAPKPEAPTTTPKETAPTTTPKKLTPTTPEKLAPTTPEKAPTTTPELAP 557
Db 500 EPAPTTTKKAPTTAPKPEAPTTTPKETAPTTTPKKLTPTTPEKLAPTTPEKAPTTTPELAP 559
QY 558 TTPPEPTPTTPPEAPTTPKAAPNTPKPEAPTTKPEAPTTKPEAPTTTPKETAPTTTPK 617
Db 560 TTPPEPTPTTPPEAPTTPKAAPNTPKPEAPTTKPEAPTTKPEAPTTTPKETAPTTTPK 619
QY 618 GTAPTTLKPEAPTTPKKAPKELAPTTTKETSTSDKAPTTPKGAPTTTSDKAPTTTPKETAPT 677
Db 620 GTAPTTLKPEAPTTPKKAPKELAPTTTKETSTSDKAPTTPKGAPTTTSDKAPTTTPKETAPT 679
QY 678 KPEAPTTPKGAPTTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKETAPT 737
Db 680 KPEAPTTPKGAPTTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKAPTTTPKETAPT 739
QY 738 PKPEAPTTPKKAPTTPEPTTSPVSTPTTKPEPTTIHKSPDESTPELSAETPKALE 797
Db 740 PKPEAPTTPKKAPTTPEPTTSPVSTPTTKPEPTTIHKSPDESTPELSAETPKALE 799
QY 798 NSPKPEPGVPTTKTAAKPEMTTAKDKTTERDLRTTPTTAAKPKMTKETATTTTEKTT 857
Db 800 NSPKPEPGVPTTKTAAKPEMTTAKDKTTERDLRTTPTTAAKPKMTKETATTTTEKTT 859
QY 858 SKITATTTQVSTTTQDTPPTPKITLKTTLAPKVTTTKKTTTTEIMNKPETAKPKDR 917
Db 860 SKITATTTQVSTTTQDTPPTPKITLKTTLAPKVTTTKKTTTTEIMNKPETAKPKDR 919
QY 918 ATNSKATTPKPKPTKAPKPTSTKPKTMPVRKPKTTPPKMTSTMPELNPTSRAE 977
Db 920 ATNSKATTPKPKPTKAPKPTSTKPKTMPVRKPKTTPPKMTSTMPELNPTSRAE 979
QY 978 AMLQTTTREPNTNPSKLVEVNPKSDAGAGETPHMLLRHVFMPEVTPDMXDYLPRVN 1037
Db 980 AMLQTTTREPNTNPSKLVEVNPKSDAGAGETPHMLLRHVFMPEVTPDMXDYLPRVN 1039
QY 1038 QGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWG 1097
Db 1040 QGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFSPSPARRITEVWG 1099
QY 1098 IPSPTDVTFTCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIWAALSTA 1157
Db 1100 IPSPTDVTFTCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIWAALSTA 1159
QY 1158 KYKNWPSVYFVKRGGSIQQYIYKQEPVQKCGRRPALNRYVYGMTQVRRRRFERAIGP 1217
Db 1160 KYKNWPSVYFVKRGGSIQQYIYKQEPVQKCGRRPALNRYVYGMTQVRRRRFERAIGP 1219

QY 1218 SQTHTIRIQSPARLAYQDKGVILNEVKVSLMRGLPNVVTSAISLPIRKPQDGYDYAF 1277
Db 1220 SQTHTIRIQSPARLAYQDKGVILNEVKVSLMRGLPNVVTSAISLPIRKPQDGYDYAF 1279
QY 1278 SKDQYVNDIVPSRTARAITTTSQTLSKVVWYVNC 1311
Db 1280 SKDQYVNDIVPSRTARAITTTSQTLSKVVWYVNC 1313
RESULT 11
US-07-757-022B-46
; Sequence 46, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Geert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-46
Query Match 96.0%; Score 6784.9; DB 4; Length 1320;
Best Local Similarity 93.3%; Pred. No. 3.9e-185;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
QY 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRATCNCNDYNCQHYWCCPDF 60
Db 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGYSRATCNCNDYNCQHYWCCPDF 60
QY 61 KRVCATLSCKGRCFESFERGRCDCDAQCKKYDKCCPDFYESFCA----- 105

Db 26 -----ELSCRGCFESFERGECDCDAQCKYDKCCDFYSFCAEVHNP'SPSPSKKAP 79
Qy 106 -----EVKDNKNR'TKKXPTPKPPVVDAG 130
Db 80 PPSGASQIKSTTKESPKNPKKTKKVIESBEITEVKDNKNR'TKKXPTPKPPVVDAG 139
Qy 131 SGLDNGDFKVTTPDSTTQHNKVSTSPKIKTAKPINRPSLPNPSDTSKETSILTUNKETT 190
Db 140 SGLDNGDFKVTTPDSTTQHNKVSTSPKIKTAKPINRPSLPNPSDTSKETSILTUNKETT 199
Qy 191 VETKETTITNKTSTDGKKTTSKETSIAKOLAPTSKVLAKPTPKAETTTKSPA 250
Db 200 VETKETTITNKTSTDGKKTTSKETSIAKOLAPTSKVLAKPTPKAETTTKSPA 259
Qy 251 LTTTPKEPTTTPKEPASITTPKEPTTTPKSAETTTKPAETTTKSPAPTTTKE 310
Db 260 LTTTPKEPTTTPKEPASITTPKEPTTTPKSAETTTKPAETTTKSPAPTTTKE 319
Qy 311 PAPTTPKEPAPTTPKEPAPTTPKSAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 370
Db 320 PAPTTPKEPAPTTPKEPAPTTPKSAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 379
Qy 371 PKEPAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 430
Db 380 PKEPAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 439
Qy 431 PTTTTSAPTTPKEPAPTTPKSAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 490
Db 440 PTTTTSAPTTPKEPAPTTPKSAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 499
Qy 491 PAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 550
Db 500 PAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 559
Qy 551 TPELAPTTPEPTTPKEPAPTTPKAAAPNTPKPAETTTKPAETTTKSPAPTTTKE 610
Db 560 TPELAPTTPEPTTPKEPAPTTPKAAAPNTPKPAETTTKPAETTTKSPAPTTTKE 619
Qy 611 TAPTTPKGTAPTTLKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 670
Db 620 TAPTTPKGTAPTTLKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 679
Qy 671 EPAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 730
Db 680 EPAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 739
Qy 731 KETAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 790
Db 740 KETAPTTPKEPAPTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 799
Qy 791 PTPKALENSPKPVGVTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 850
Db 800 PTPKALENSPKPVGVTTPKPAETTTKPAETTTKSPAPTTTKEPAPTTPKEPTT 859
Qy 851 TTEKTTESKLTATTTQVSTTTQDTPPKIITLTKTTLAPKVTTKTITTEIMNKEE 910
Db 860 TTEKTTESKLTATTTQVSTTTQDTPPKIITLTKTTLAPKVTTKTITTEIMNKEE 919
Qy 911 TAKPKDRATNSKATTPKPKETPKKPTSTKPKETMPRVKPKTTPPRKMTSTMPBLN 970
Db 920 TAKPKDRATNSKATTPKPKETPKKPTSTKPKETMPRVKPKTTPPRKMTSTMPBLN 979
Qy 971 PTSRIAEAMLQTTTRPNQTPNSKLVEVNPKEDEAGAGETPHMLLRPHVFWPVTDPMD 1030
Db 980 PTSRIAEAMLQTTTRPNQTPNSKLVEVNPKEDEAGAGETPHMLLRPHVFWPVTDPMD 1039
Qy 1031 YLPRVNPNGIILNPMLSDETNI'CNCKPVDGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1090
Db 1040 YLPRVNPNGIILNPMLSDETNI'CNCKPVDGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1099
Qy 1091 RITEVWGI'PSIDT'VTRCNCEGKTF'FFKDSQYWR'FNDIKDAGY'PKDIFKGFGLTQOI 1150
Db 1100 RITEVWGI'PSIDT'VTRCNCEGKTF'FFKDSQYWR'FNDIKDAGY'PKDIFKGFGLTQOI 1159

Qy 1151 VAALSTAKYKNWPESVYFFKRGSGSIQQYIYKQEPVQKCPGRPALNYPVYGEMTQVRRRR 1210
Db 1160 VAALSTAKYKNWPESVYFFKRGSGSIQQYIYKQEPVQKCPGRPALNYPVYGEMTQVRRRR 1219
Qy 1211 FERAIGPSQTHTIRIOYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPTRKPD 1270
Db 1220 FERAIGPSQTHTIRIOYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPTRKPD 1279
Qy 1271 GYDYAFSKDOQYNTDVPSTARATITTSQGTL'SKWNVNC 1311
Db 1280 GYDYAFSKDOQYNTDVPSTARATITTSQGTL'SKWNVNC 1320

RESULT 12

US-07-757-022B-60
; Sequence 60, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-60

Query Match 96.0%; Score 6784.9; DB 4; Length 1320;
Best Local Similarity 93.3%; Pred. No. 3.9e-185;
Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
Qy 1 MAWKTLPIYLILLISLVFVIQQVSSQDLSSCAGRCGEGYSRDTACNCDYNCQHMECCPDF 60

Db 1 NAWKTLPIYLLGLLSVFIQVSSQ----- 25
QY 61 KRVCETAELSKGRGCFESPERGECDDAQCKYDKCCPDYSEFCA----- 105
Db 26 -----ELSKGRGCFESPERGECDDAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
QY 106 -----EVKDNKKNRITKKKPTPKPPVDEAG 130
Db 80 PPGASQTIKSTTKRSPKPNKKTKKVISEETEVKDNKKNRITKKKPTPKPPVDEAG 139
QY 131 SGLDNGDFKVTTPDTSSTOHNKUSTSPKIIITAKINPRAPSLPPNSDTSKETSILVKNKETT 190
Db 140 SGLDNGDFKVTTPDTSSTOHNKUSTSPKIIITAKINPRAPSLPPNSDTSKETSILVKNKETT 199
QY 191 VETKETTITNKQTSDDGKEKITSKETSQTSIKETSADKLAPTSKVLAKPTPKAETTTKGA 250
Db 200 VETKETTITNKQTSDDGKEKITSKETSQTSIKETSADKLAPTSKVLAKPTPKAETTTKGA 259
QY 251 LITPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 310
Db 260 LITPKETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 319
QY 311 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTP 370
Db 320 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTP 379
QY 371 PKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTP 430
Db 380 PKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTP 439
QY 431 PTTTKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 490
Db 440 PTTTKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 499
QY 491 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 550
Db 500 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 559
QY 551 TPEELAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 610
Db 560 TPEELAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 619
QY 611 TAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 670
Db 620 TAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 679
QY 671 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 730
Db 680 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 739
QY 731 KETAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 790
Db 740 KETAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 799
QY 791 PTPKALNSPKBPVPTTKTAAKPEMTTAAKKTTERDLRTTPTTAAKMTKETAT 850
Db 800 PTPKALNSPKBPVPTTKTAAKPEMTTAAKKTTERDLRTTPTTAAKMTKETAT 859
QY 851 TTEKTTESKITAATTQVSTTODTTPKIIITLKTTLAPKVTTTKTITTEIMNKPEE 910
Db 860 TTEKTTESKITAATTQVSTTODTTPKIIITLKTTLAPKVTTTKTITTEIMNKPEE 919
QY 911 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMPRVKPKTTPPKMTSTMPBLN 970
Db 920 TAKPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMPRVKPKTTPPKMTSTMPBLN 979
QY 971 PTSRIABAMLQTTTPNQTSPNSKLVENPKSDAGGAGETPHMLLRPHVFMPEVTPDMD 1030
Db 980 PTSRIABAMLQTTTPNQTSPNSKLVENPKSDAGGAGETPHMLLRPHVFMPEVTPDMD 1039
QY 1031 YLPRVPGOGIILNMLSDETNICNGKPVGDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1090

Db 1040 YLPRVPGOGIILNMLSDETNICNGKPVGDGLTTLRNGTLVAFRGHYFWMLSPFSPSPAR 1099
QY 1091 RITEVWGIPSPIDTVTRCNCCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGOI 1150
Db 1100 RITEVWGIPSPIDTVTRCNCCEGKTFPFKDSQYWRFTNDIKDAGYPKPIFKGFGGLTGOI 1159
QY 1151 VAALSTAKYKNWPSVVFVKRGGSIQOYIYKQBPVQKCPGRRPALNYPVYGEMTQVRRRR 1210
Db 1160 VAALSTAKYKNWPSVVFVKRGGSIQOYIYKQBPVQKCPGRRPALNYPVYGEMTQVRRRR 1219
QY 1211 FERAIGSQHTIRIQSPARLAYODKGVHLNPKVSLWRGLPNNVTSAISLNPRIKPD 1270
Db 1220 FERAIGSQHTIRIQSPARLAYODKGVHLNPKVSLWRGLPNNVTSAISLNPRIKPD 1279
QY 1271 GYDYAFSKOQYNNIDVPSRTARAITTRSGQTLSKWWYNCP 1311
Db 1280 GYDYAFSKOQYNNIDVPSRTARAITTRSGQTLSKWWYNCP 1320

RESULT 13
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-52

| | | | |
|----|------|--|------|
| QY | 311 | PAPTTPKEPAPTTTKBPAPTTTKSAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT | 370 |
| Db | 320 | PAPTTPKEPAPTTTKBPAPTTTKSAPTTTKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTT | 379 |
| QY | 371 | PKEPAPTTKEPAPTTTPKEPAPTA PKKAPATTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE | 430 |
| Db | 380 | PKEPAPTTKEPAPTTTPKEPAPTA PKKAPATTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE | 439 |
| QY | 431 | PTTTKSAPTTTKBPAPTTTKSAPTTTPKESPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKE | 490 |
| Db | 440 | PTTTKSAPTTTKBPAPTTTKSAPTTTPKESPTTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKE | 499 |
| QY | 491 | PAPTTPKEPAPTTTKKAPATA PKBPAPTTTPKEPAPTTTPKCLPTTPKEPAPTTTPKEPAPTT | 550 |
| Db | 500 | PAPTTPKEPAPTTTKKAPATTTPKEPAPTTTPKETA PTTPKLPTTPKEPAPTTTPKEPAPTT | 559 |
| QY | 551 | TPKEELAPTTPEBPTPTTPPEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE | 610 |
| Db | 560 | TPKEELAPTTPEBPTPTTPPEPAPTTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE | 619 |
| QY | 611 | TAPTTPKGTAPTTLKEPAPTTTPKEPAPKELAPTTTPKEPTSTSDKAPATTTPKGTAPTTTPK | 670 |
| Db | 620 | TAPTTPKGTAPTTLKEPAPTTTPKGPAPKELAPTTTPKEPTSTSDKAPATTTPKGTAPTTTPK | 679 |
| QY | 671 | EPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKGPAPKELAPTTTPKGTSTTSXKAPATTTP | 730 |
| Db | 680 | EPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTTPKGPAPKELAPTTTPKGTSTTSXKAPATTTP | 739 |
| QY | 731 | KETAPTTPEKAPATTTPKAPAPTTTPETTPPTTSEVSTPTTKAPTTHKSPDESTPELSAE | 790 |
| Db | 740 | KETAPTTPEKAPATTTPKAPAPTTTPETTPPTTSEVSTPTTKAPTTHKSPDESTPELSAE | 799 |
| QY | 791 | PTPKALENSPKBPVPTTKTPAATKPEMTTAKOKTTERDLRTPETTTAAPKMTKETAT | 850 |
| Db | 800 | PTPKALENSPKBPVPTTKTPAATKPEMTTAKOKTTERDLRTPETTTAAPKMTKETAT | 859 |
| QY | 851 | TTEKTTESKIATTTTQVTSITTTQDTPPKITTLTKTTLAPKVTTKKTIITTTTEIMNKPEE | 910 |
| Db | 860 | TTEKTTESKIATTTTQVTSITTTQDTPPKITTLTKTTLAPKVTTKKTIITTTTEIMNKPEE | 919 |
| QY | 911 | TAKPKDRATNSKATTPPKQPKETKAPKPTSTKPKTWPVRVKPPTTPPKMTSTMPBLN | 970 |
| Db | 920 | TAKPKDRATNSKATTPPKQPKETKAPKPTSTKPKTWPVRVKPPTTPPKMTSTMPBLN | 979 |
| QY | 971 | PTSRISAEMLQTTTPRPNQTPNSKULVENPKSEDAGGAETPHMLLPHVFMPEVITDMD | 1030 |
| Db | 980 | PTSRISAEMLQTTTPRPNQTPNSKULVENPKSEDAGGAETPHMLLPHVFMPEVITDMD | 1039 |
| QY | 1031 | YLPVRPNQGI IINPMLSDETNICNGKVPDGLTTLRNGTLVAFRGHYFMMI.SPPSPSPAR | 1090 |
| Db | 1040 | YLPVRPNQGI IINPMLSDETNICNGKVPDGLTTLRNGTLVAFRGHYFMMI.SPPSPSPAR | 1099 |
| QY | 1091 | RI TEVWGI PGPIDTVTRCNCCEGKTFFPKDSQYWRFTNDIKDAGYKXPIFKFGGLTGQI | 1150 |
| Db | 1100 | RI TEVWGI PGPIDTVTRCNCCEGKTFFPKDSQYWRFTNDIKDAGYKXPIFKFGGLTGQI | 1159 |
| QY | 1151 | VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQBQVKQCPGRRALNYPVYGETQVRRRR | 1210 |
| Db | 1160 | VAALSTAKYKNWPESVYFFKRGSGIOQYIYKQBQVKQCPGRRALNYPVYGETQVRRRR | 1219 |
| QY | 1211 | FERAIGSQHTTIRIQYSPARLAYQDKGVILHNEVKYSILWRGLPNVVTSAISLPNTRKPD | 1270 |
| Db | 1220 | FERAIGSQHTTIRIQYSPARLAYQDKGVILHNEVKYSILWRGLPNVVTSAISLPNTRKPD | 1279 |
| QY | 1271 | GXDYVAFSKDQYINIDVPSRTABAITTRSQTLTSKQVWYNCP | 1311 |
| Db | 1280 | GXDYVAFSKDQYINIDVPSRTABAITTRSQTLTSKQVWYNCP | 1320 |

RESULT 15

US-07-757-022B-104

; Sequence 104, Application US/07757022B

Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 104:

SEQUENCE CHARACTERISTICS:

LENGTH: 1140 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-104

| Query Match | Score | DB 4 | Length |
|-------------|-------|--------|--------|
| Best Local | 79.3% | 5603.7 | 1140 |
| Global | 81.8% | 5603.7 | 1140 |

| | Matches | 1047; | Conservative | 0; | Mismatches | 135E-151; | Indels | 93; | Gaps | 1; |
|----|---------|---|--------------|----|------------|-----------|--------|-----|------|----|
| Qy | 1 | MAWKTLPIVILLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHWECCPDF | 60 | | | | | | | |
| Db | 1 | MAWKTLPIVILLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHWECCPDF | 60 | | | | | | | |
| Qy | 61 | KRVCTAELSCKRCFESFERGREGCDCAOCKYDKCCPDYESFCAE- | 106 | | | | | | | |
| Db | 61 | KRVCTAELSCKRCFESFERGREGCDCAOCKYDKCCPDYESFCAEVHNPTSPSSKKAP | 120 | | | | | | | |
| Qy | 107 | ----- | 106 | | | | | | | |
| Db | 121 | PPSGASQTIKSTKSPKPNKKTKVIESEBITTEHSVSENQBSSSSSSSSSSTIW | 180 | | | | | | | |
| Qy | 107 | -----VKDNKNRTRKKETPKPPVVDEAGSLDNGDFKVTTPDTST | 147 | | | | | | | |
| Db | 181 | KIKSKNSAANRELQKLLVKDNKNRTRKKETPKPPVVDEAGSLDNGDFKVTTPDTST | 240 | | | | | | | |
| Qy | 148 | TOHNKVSTSPKITTAKPINPRSLPNSDTSKETSLTWNKETTVEKTTTTNKOTSTDG | 207 | | | | | | | |

seq1-h.rai

Wed Oct 13 12:38:28 2004

```

Db      241  TQHNKVSTSPKITTAKP.INPRPSLPDNDSTSKETSLTVNKETTVETKETTINKQSTSDG 300
Qy      208  KEKTTSAKETOSIEKTSKDLAPTSKVLAKPTPKAETTTKGFPALTPKPEPTPTTPKBPAS 267
Db      301  KEKTTSAKETOSIEKTSKDLAPTSKVLAKPTPKAETTTKGFPALTPKPEPTPTTPKBPAS 360
Qy      268  TTPKEPTTTTKSAPTTKBPAPTTKGAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 327
Db      361  TTPKEPTTTTKSAPTTKBPAPTTKGAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 420
Qy      328  APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 387
Db      421  APTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 480
Qy      388  BPAPTAPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 447
Db      481  BPAPTAPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 540
Qy      448  TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 507
Db      541  TTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 600
Qy      508  APTAPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 567
Db      601  APTAPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 660
Qy      568  PEPAPPTPKAAAPNTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 627
Db      661  PEPAPPTPKAAAPNTPKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 720
Qy      628  APTPKKAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 687
Db      721  APTPKKAPKELAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 780
Qy      688  TAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 747
Db      781  TAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBP 840
Qy      748  KPAPTTPETPPPTTSEVSTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 807
Db      841  KPAPTTPETPPPTTSEVSTPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTK 900
Qy      808  TKTPAATKPEMTTAKDKITTERDLRTTPTTAAAPKMTKETATTTTEKTTESKITATTQV 867
Db      901  TKTPAATKPEMTTAKDKITTERDLRTTPTTAAAPKMTKETATTTTEKTTESKITATTQV 960
Qy      868  TSTTTQDTPPKITTLKTTILAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATPK 927
Db      961  TSTTTQDTPPKITTLKTTILAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATPK 1020
Qy      928  PQKETKAPKPTSTKPKMTMPRVKPKTTPTPRKMTSTMPBELNPTSIAEAMLOTTTREN 987
Db      1021  PQKETKAPKPTSTKPKMTMPRVKPKTTPTPRKMTSTMPBELNPTSIAEAMLOTTTREN 1080
Qy      988  QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1047
Db      1081  QTPNSKLVNPKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1140

```

Search completed: October 13, 2004, 11:59:29
Job time : 29.7891 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 133.379 Seconds
(without alignments)
3171.696 Million cell updates/sec

Title: SEQ1-H
Perfect score: 7064
Sequence: 1 MAWKTLPIYLLLLSVFIQ.....ARAITRSGQTLSKVWNCP 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------------|
| 1 | 7064 | 100.0 | 1311 | 13 | US-10-124-557-42 |
| 2 | 7049.7 | 99.8 | 1354 | 13 | US-10-124-557-48 |
| 3 | 7049 | 99.8 | 1361 | 13 | US-10-124-557-40 |
| 4 | 7044.7 | 99.7 | 1404 | 9 | US-09-802-207-30 |
| 5 | 7044.7 | 99.7 | 1404 | 11 | US-09-897-188-1 |
| 6 | 7044.7 | 99.7 | 1404 | 13 | US-10-124-557-2 |
| 7 | 7044.7 | 99.7 | 1404 | 13 | US-10-124-557-62 |
| 8 | 6811.3 | 96.4 | 1314 | 13 | US-10-124-557-50 |
| 9 | 6799.9 | 96.3 | 1270 | 13 | US-10-124-557-44 |
| 10 | 6785.6 | 96.1 | 1313 | 13 | US-10-124-557-142 |
| 11 | 6784.9 | 96.0 | 1320 | 13 | US-10-124-557-46 |
| 12 | 6784.9 | 96.0 | 1320 | 13 | US-10-124-557-60 |
| 13 | 6780.6 | 96.0 | 1363 | 13 | US-10-124-557-52 |
| 14 | 5603.7 | 79.3 | 1140 | 13 | US-10-124-557-104 |

```

15 5500.7 77.9 1038 13 US-10-124-557-74
16 5472.6 77.5 1022 13 US-10-124-557-84
17 5344.6 75.7 1049 13 US-10-124-557-58
18 5011 70.9 941 13 US-10-124-557-14
19 3680.5 52.1 792 9 US-09-802-207-27
20 2850.9 40.4 538 14 US-10-038-694-3
21 2057.8 29.1 463 13 US-10-124-557-54
22 1806.8 25.6 423 13 US-10-124-557-66
23 1793.7 25.4 422 13 US-10-124-557-68
24 1723.1 24.4 372 13 US-10-124-557-64
25 1714.7 24.3 401 9 US-09-802-207-29
26 1398.2 19.8 5179 9 US-09-922-217-1068
27 1398.2 19.8 5179 9 US-09-833-263-1068
28 1398.2 19.8 5179 9 US-10-025-380-1068
29 1398.2 19.8 5179 16 US-10-734-564-121
30 1307.1 18.5 292 16 US-10-468-910-4
31 1219.6 17.3 237 13 US-10-124-557-72
32 1108.5 15.7 296 13 US-10-124-557-70
33 1025 14.5 188 14 US-10-038-694-2
34 969.9 13.7 1460 14 US-10-295-027-428
35 956.2 13.5 1367 9 US-09-801-368-108
36 951.7 13.5 3507 14 US-10-369-493-5784
37 945 13.4 6642 14 US-10-369-493-5013
38 941.5 13.3 19723 15 US-10-084-846A-5
39 931.3 13.2 1325 9 US-09-864-761-35612
40 884 12.5 157 13 US-10-124-557-102
41 884 12.5 157 13 US-10-124-557-114
42 879.4 12.4 5935 14 US-10-243-243A-8
43 877.1 12.4 5877 14 US-10-142-515-11
44 874.2 12.4 22152 16 US-10-715-066-5
45 873.5 12.4 2090 16 US-10-408-765A-2318

```

ALIGNMENTS

RESULT 1

```

US-10-124-557-42
; Sequence 42, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

```

ATTORNEY/AGENT INFORMATION:
NAME: Ceerit, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

Query Match 100.0%; Score 7064; DB 13; Length 1311;
Best Local Similarity 100.0%; Pred. No. 3e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNCOHYMECCPDF 60
QY 61 KRVCYTAELSCGRCFESFERGECDCDAQCKYDKCCPDYESFCAEVKONKNTKKPT 120
Db 61 KRVCYTAELSCGRCFESFERGECDCDAQCKYDKCCPDYESFCAEVKONKNTKKPT 120
QY 121 KRPVVDVDEAGSLDNGDFKVTTPDSTQHNKYSTSPKITTAKPINRPSLPNSDTSKE 180
Db 121 KRPVVDVDEAGSLDNGDFKVTTPDSTQHNKYSTSPKITTAKPINRPSLPNSDTSKE 180
QY 181 TSLVVKETIVETKETTNNKQISTDGKETTSAKETQSIEKTSKADLAFTSKVLAKPTP 240
Db 181 TSLVVKETIVETKETTNNKQISTDGKETTSAKETQSIEKTSKADLAFTSKVLAKPTP 240
QY 241 KAETTTKGPALTPKBPPTTPKEBPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTP 300
Db 241 KAETTTKGPALTPKBPPTTPKEBPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTP 300
QY 301 KEPAPTTTKEPATTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 360
Db 301 KEPAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 360
QY 361 TTPKEPTPTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKE 420
Db 361 TTPKEPTPTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKE 420
QY 421 PSPTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTT 480
Db 421 PSPTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTT 480
QY 481 KKPAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 540
Db 481 KKPAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 540
QY 541 PTTPEKAPTTTPEELAPTTPEEPPTTPEEBAPTTTPEEBAPTTTPEEBAPTTTPEEBAPTT 600
Db 541 PTTPEKAPTTTPEELAPTTPEEPPTTPEEBAPTTTPEEBAPTTTPEEBAPTTTPEEBAPTT 600
QY 601 KEPAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 660
Db 601 KEPAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 660
QY 661 PKGTAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 720
Db 661 PKGTAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 720
QY 721 TSDKAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 780
Db 721 TSDKAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT 780
QY 781 DESTPELSAEPTPKALENSPKPGVPTTKTAAATKPEMTTTAKDKTTERDLATTPETTTA 840

Db 781 DESTPELSAEPTPKALENSPKPGVPTTKTAAATKPEMTTTAKDKTTERDLATTPETTTA 840
QY 841 APKMTKETATTTTEKTTESKITATTTQVTSSTTQDTPPKITTLTKTTTLAPKVTTKKTTIT 900
Db 841 APKMTKETATTTTEKTTESKITATTTQVTSSTTQDTPPKITTLTKTTTLAPKVTTKKTTIT 900
QY 901 TTEIMNKPEETAKPKDRATNSKATTPKQKPKTKAPKPKTSTTKPKTMFVRVKPKTTPTPR 960
Db 901 TTEIMNKPEETAKPKDRATNSKATTPKQKPKTKAPKPKTSTTKPKTMFVRVKPKTTPTPR 960
QY 961 KMTSTMPELNPTSRIAEAMLQTTTPNQTNPNSKLVENPKSDEADAGAGETPHMLLRPHV 1020
Db 961 KMTSTMPELNPTSRIAEAMLQTTTPNQTNPNSKLVENPKSDEADAGAGETPHMLLRPHV 1020
QY 1021 FMPEVTPDMYLPFRVNPQGIILNPMLSDETNCNGKPDVGLTTLRNGTLVAFRGHYFWML 1080
Db 1021 FMPEVTPDMYLPFRVNPQGIILNPMLSDETNCNGKPDVGLTTLRNGTLVAFRGHYFWML 1080
QY 1081 SPFSPSPARRITEVWGIPSPIDTVTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIF 1140
Db 1081 SPFSPSPARRITEVWGIPSPIDTVTRCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIF 1140
QY 1141 KGFGGLTGQIIVAALSTAKYNWPEVYFFKRGSSIIOYIYKQEPVQKCGRRPALNYPVY 1200
Db 1141 KGFGGLTGQIIVAALSTAKYNWPEVYFFKRGSSIIOYIYKQEPVQKCGRRPALNYPVY 1200
QY 1201 GEMTQVRRRRFERAIGPSQTHIRIOYSPARLAYQKGLHNEVKVSIILWRGLPNVVTSA 1260
Db 1201 GEMTQVRRRRFERAIGPSQTHIRIOYSPARLAYQKGLHNEVKVSIILWRGLPNVVTSA 1260
QY 1261 ISLPNIRKPDGYDYAFSKDQYNNIDVPSRTARAITTRSGQTLSKVWYNCP 1311
Db 1261 ISLPNIRKPDGYDYAFSKDQYNNIDVPSRTARAITTRSGQTLSKVWYNCP 1311

RESULT 2

US-10-124-557-48
; Sequence 48, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserit, Luan
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-10-124-557-48

Query Match          99.8%; Score 7049.7; DB 13; Length 1354;
Best Local Similarity 96.8%; Pred. No. 6.9e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIYLLLSVFEVIQQVSSQDLSSCAGRCGEGYSRDATCNCDCYNCOHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFEVIQQVSSQDLSSCAGRCGEGYSRDATCNCDCYNCOHYMECCPDF 60
QY 61 KRVTAE LSCGRCFESFERGREGCDCAQCKYDKCCPDYESFCAE----- 106
DB 61 KRVTAE LSCGRCFESFERGREGCDCAQCKYDKCCPDYESFCAEHSHSVSENCBSSSS 120
QY 107 -----VKDNKKQRTKKKPTKPPVDEAGSLDNGD 137
DB 121 SSSSSSTIWKIYSSKNSAANRELQKKLVKDNKKQRTKKKPTKPPVDEAGSLDNGD 180
QY 138 FKVTPTDSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETT 197
DB 181 FKVTPTDSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETT 240
QY 198 TTNKQISTDCKEKTSAKETQSTIEKTSKADLAPTSKVLAKPTPKAEITTKGPAITTPKEP 257
DB 241 TTNKQISTDCKEKTSAKETQSTIEKTSKADLAPTSKVLAKPTPKAEITTKGPAITTPKEP 300
QY 258 TPTPKPEASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTPK 317
DB 301 TPTPKPEASTTPKEPTPTTIKSAPTTPKEPATTTKSAPTTPKBPATTTKBPATTPK 360
QY 318 EPAPTTKBPATTTKSAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 377
DB 361 EPAPTTKBPATTTKSAPTTPKEPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 420
QY 378 TKEPATTTKBPATAPKAPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 437
DB 421 TKEPATTTKBPATAPKAPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 480
QY 438 PTTTKBPATTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 497
DB 481 PTTTKBPATTTKSAPTTPKBPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 540
QY 498 EPAPTTKBPATAPKAPATTPKBPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 557
DB 541 EPAPTTKBPATAPKAPATTPKBPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 600
QY 558 TTBEPPTTTBEPATTPKAAAPNPKBPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 617
DB 601 TTBEPPTTTBEPATTPKAAAPNPKBPATTPKBPATTPKBPATTPKBPATTTKBPATTTKSA 660
QY 618 GTAPTTLKBPATTPKBPAPKELAPTTKBPATTTSDKPAITTPKBPATTPKBPATTTK 677
DB 661 GTAPTTLKBPATTPKBPAPKELAPTTKBPATTTSDKPAITTPKBPATTPKBPATTTK 720
QY 678 KEPATTPKBPATTPKBPAPKELAPTTKBPATTTSDKPAITTPKBPATTPKBPATTTK 737
DB 721 KEPATTPKBPATTPKBPAPKELAPTTKBPATTTSDKPAITTPKBPATTPKBPATTTK 780
QY 738 PKEPATTPKBPATTPPTPPPTTSEVSTPTTTKEPTTIHKSPOBSESTPELSAEPKPALE 797

```

RESULT 3

```

US-10-124-557-40
; Sequence 40, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
;           Clark, Stephen C.
;           Jacobs, Kenneth
;           Hewick, Rodney M.
;           Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990

```



```

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
QY 61 KRVCCTAELSCKGRCFESPERGECDCDAQCKKYDKCCPDYESFCAE----- 106
Db 61 KRVCCTAELSCKGRCFESPERGECDCDAQCKKYDKCCPDYESFCAEVENHPTSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEEITEHSVSENQESSSSSSSSSIW 180
QY 107 -----VKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 147
Db 181 KIKSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 240
QY 148 TQHNKVSIPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQSTSDG 207
Db 241 TQHNKVSIPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQSTSDG 300
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 267
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 360
QY 268 TTPKEPTPTTIKSAPITPKAPPTTKSAPITPKAPPTTKGPAPTTTKGPAPTTTKGP 327
Db 361 TTPKEPTPTTIKSAPITPKAPPTTKSAPITPKAPPTTKGPAPTTTKGPAPTTTKGP 420
QY 328 APITTKSAPITPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 387
Db 421 APITTKSAPITPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 480
QY 388 EPAPTAPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 447
Db 481 EPAPTAPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 540
QY 448 TTKSAPITPKESPITTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGP 507
Db 541 TTKSAPITPKESPITTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 600
QY 508 APAPTAPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 567
Db 601 APAPTAPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 660
QY 568 PEPAPTTPKAAAPNTPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 627
Db 661 PEPAPTTPKAAAPNTPKAPPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 720
QY 628 APPTPKAPKELAPITTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGP 687
Db 721 APPTPKAPKELAPITTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 780
QY 688 TAPPTLKAPPTTKKAPKELAPITTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 747
Db 781 TAPPTLKAPPTTKKAPKELAPITTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 840
QY 748 KPAPTTPPTTSEVSTPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGP 807
Db 841 KPAPTTPPTTSEVSTPTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 900
QY 808 TKTPAATKEMTTTAKDITRDLRTTPTTAAKMTKETATTTKTESKITATTTQV 867
Db 901 TKTPAATKEMTTTAKDITRDLRTTPTTAAKMTKETATTTKTESKITATTTQV 960
QY 868 TSTTTQDITTPPKITTLTKITTLAPKVTTKKITTTTTEIMNKBEETAKPKORATNSKATTPK 927
Db 961 TSTTTQDITTPPKITTLTKITTLAPKVTTKKITTTTTEIMNKBEETAKPKORATNSKATTPK 1020
QY 928 POKPTAPKPKPTSTIKKPKTMRVRKPKTTPPKMTSTMPBLNPTSRFAEAMLOTTTPRN 987
Db 1021 POKPTAPKPKPTSTIKKPKTMRVRKPKTTPPKMTSTMPBLNPTSRFAEAMLOTTTPRN 1080
QY 988 QTPNSKLVENVNPKBEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1047

```

```

Db 1081 QTPNSKLVENVNPKBEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140
QY 1048 DETNICNGKPGVDGLTTLRNGTLVAFRGHYFWMLSPESSPARRITEVWGPSPIDTVPFT 1107
Db 1141 DETNICNGKPGVDGLTTLRNGTLVAFRGHYFWMLSPESSPARRITEVWGPSPIDTVPFT 1200
QY 1108 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1167
Db 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIIVAALSTAKYKNWPESVY 1260
QY 1168 FFKEGGSIQQYIYKQEPVQKCPGRPALNPVYGEQVRRRRPERRAIGPSQTHIRIQY 1227
Db 1261 FFKEGGSIQQYIYKQEPVQKCPGRPALNPVYGEQVRRRRPERRAIGPSQTHIRIQY 1320
QY 1228 SPARLAYQDKGVHLNHEVKVSIILWRGLPNVVTSAISLNNIRKPDGYDYAFSKOQYINIDV 1287
Db 1321 SPARLAYQDKGVHLNHEVKVSIILWRGLPNVVTSAISLNNIRKPDGYDYAFSKOQYINIDV 1380
QY 1288 PSRTARAITTRSGQTLSSKWYNCP 1311
Db 1381 PSRTARAITTRSGQTLSSKWYNCP 1404

```

RESULT 5

```

US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1

```

```

Query Match          99.7%; Score 7044.7; DB 11; Length 1404;
Best Local Similarity 93.4%; Pred. No. 9.6e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

```

```

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
QY 61 KRVCCTAELSCKGRCFESPERGECDCDAQCKKYDKCCPDYESFCAE----- 106
Db 61 KRVCCTAELSCKGRCFESPERGECDCDAQCKKYDKCCPDYESFCAEVENHPTSPSSKKAP 120
QY 107 ----- 106
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIIESEEITEHSVSENQESSSSSSSSSIW 180
QY 107 -----VKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 147
Db 181 KIKSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDST 240
QY 148 TQHNKVSIPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQSTSDG 207
Db 241 TQHNKVSIPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTINKQSTSDG 300
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 267
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 360

```


QY 148 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNNKQSTG 207
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNNKQSTG 300
QY 208 KEKTSKAKETQSEKTSKAKOLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 267
DB 301 KEKTSKAKETQSEKTSKAKOLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 360
QY 268 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKP 327
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKP 420
QY 328 APPTTKSAPTTPKEPAPTTPKKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTKPAPTTPK 387
DB 421 APPTTKSAPTTPKEPAPTTPKKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTKPAPTTPK 480
QY 388 EPAPTAPKAPPTTPKEPAPTTPKKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTKSAPTTPKEPAPTT 447
DB 481 EPAPTAPKAPPTTPKEPAPTTPKKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTKSAPTTPKEPAPTT 540
QY 448 TTKSAPTTPKEPAPTTTPKEPAPTTPKKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTKSAPTTPKEPAPTT 507
DB 541 TTKSAPTTPKEPAPTTTPKEPAPTTPKKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTKPAPTTTKP 600
QY 508 APAPTAPKEPAPTTPKETAPTTTPKKTTPPTPEKAPTTTPKEPAPTTTPBELAPTTPEEPTPTT 567
DB 601 APAPTAPKEPAPTTPKETAPTTTPKKTTPPTPEKAPTTTPKEPAPTTTPBELAPTTPEEPTPTT 660
QY 568 PREPAPTTTPKAAAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGPATTTLKPE 627
DB 661 PREPAPTTTPKAAAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGPATTTLKPE 720
QY 628 APPTPKKAPKELAPTTTKETSTSDKAPITPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGP 687
DB 721 APPTPKKAPKELAPTTTKETSTSDKAPITPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGP 780
QY 688 TAPPTLKBPAPTTPKKAPKELAPTTTKGTSTSDKAPITPKGTAPTTTPKEPAPTTTPKEPAPTTTPK 747
DB 781 TAPPTLKBPAPTTPKKAPKELAPTTTKGTSTSDKAPITPKGTAPTTTPKEPAPTTTPKEPAPTTTPK 840
QY 748 KPAPTTTPPTPTTSEVSTPTTKPTTHKSPDESTPELSAEPPTPKALESPKPGVPT 807
DB 841 KPAPTTTPPTPTTSEVSTPTTKPTTHKSPDESTPELSAEPPTPKALESPKPGVPT 900
QY 808 TKTPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTTKTESKITATTTQV 867
DB 901 TKTPAATKPEMTTAKDKTTERDLATTPETTTAAPKMTKETATTTKTESKITATTTQV 960
QY 868 TSTTTQDTPPKITTLTKTTTLAPKVTTTKITITTEIMNKPDEATAKPKDRATNSKATTPK 927
DB 961 TSTTTQDTPPKITTLTKTTTLAPKVTTTKITITTEIMNKPDEATAKPKDRATNSKATTPK 1020
QY 928 POKPTKAPKKTSTKPKTMRVRPKTTPPRKMTSTWDELNPTSRIAEAMLOTTTRPN 987
DB 1021 POKPTKAPKKTSTKPKTMRVRPKTTPPRKMTSTWDELNPTSRIAEAMLOTTTRPN 1080
QY 988 QTPNSKLVEVPKSEDAGAGETPHMLLRHVFMEVPTPMDVLPVFNQGIINPMLS 1047
DB 1081 QTPNSKLVEVPKSEDAGAGETPHMLLRHVFMEVPTPMDVLPVFNQGIINPMLS 1140
QY 1048 DETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITEVWGIPSPDITVFT 1107
DB 1141 DETNINCNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPPFPSPARRITEVWGIPSPDITVFT 1200
QY 1108 RCNCEGKTFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIIVAALSTAKYKNWPSVY 1167
DB 1201 RCNCEGKTFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIIVAALSTAKYKNWPSVY 1260
QY 1168 FFKGSGSIQQIYIKQEPVOKPCGRRPALNYPVYGEMTQVRRRFRERAIQPSQTHIRIQY 1227
DB 1261 FFKGSGSIQQIYIKQEPVOKPCGRRPALNYPVYGEMTQVRRRFRERAIQPSQTHIRIQY 1320

QY 1228 SPARLAYQDGVGLHNEVKVSIILMRGLPNVVTSAISLPIRKPDCGYDYAFSKDQVYNIDV 1287
DB 1321 SPARLAYQDGVGLHNEVKVSIILMRGLPNVVTSAISLPIRKPDCGYDYAFSKDQVYNIDV 1380
QY 1288 PSRTARAITTRSGQTILSKWYNCP 1311
DB 1381 PSRTARAITTRSGQTILSKWYNCP 1404
RESULT 7
US-10-124-557-62
; Sequence 62, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine C.
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csert, Luana
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
US-10-124-557-62

Query Match 99.7%; Score 7044.7; DB 13; Length 1404;
Best Local Similarity 93.4%; Pred. No. 9,6e-163;
Matches 1311; Conservative 0; Mismatches 0; Indels 93; Gaps 1;
QY 1 MAWKTLPIYLLILLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLILLLSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KRVCATLSCKGRCFESFERGRCDDCAQCKYDKCCPDYSESCAE----- 106
DB 61 KRVCATLSCKGRCFESFERGRCDDCAQCKYDKCCPDYSESCAEVHNFTSPSSKAP 120

QY 1108 RCNCEGKTFPPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1167
DB 1201 RCNCEGKTFPPKDSQYWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1260
QY 1168 FFKRGSGSIQQYIKQEPVOKCPGRRPALNYPVYGMTQVRRRRFERRAIGPSQTHIRIQY 1227
DB 1261 FFKRGSGSIQQYIKQEPVOKCPGRRPALNYPVYGMTQVRRRRFERRAIGPSQTHIRIQY 1320
QY 1228 SPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKPDPGYDYAFSKDQYVNDV 1287
DB 1321 SPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPIRKPDPGYDYAFSKDQYVNDV 1380
QY 1288 PSRTARAITTRSGQTLISKVWYNCP 1311
DB 1381 PSRTARAITTRSGQTLISKVWYNCP 1404
RESULT 8
US-10-124-557-50
; Sequence 50, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Geener, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

QY 107 ----- 106
DB 121 PPSGASQIISKSTTKSPKPPNKKTKKVIIESEIITEHVSVENQSSSSSSSSSSSIW 180
QY 107 -----VKDKKRTKKKPTPKPPVVDGAGSLDNGDFKVTPTST 147
DB 181 KIKSSKNSAANRELQKLVKDKNKKRTKKPTPKPPVVDGAGSLDNGDFKVTPTST 240
QY 148 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETTVETKETTNNKOTSTDG 207
DB 241 TOHNVKSTSPKITTAKPINRPSLPNSDTSKETSITVNNKETTVETKETTNNKOTSTDG 300
QY 208 KEKTTSAKETQSIKTSKDLAPTCKVLAKPTPKAETTTKGPALTTPKPEPTTTPKEPAS 267
DB 301 KEKTTSAKETQSIKTSKDLAPTCKVLAKPTPKAETTTKGPALTTPKPEPTTTPKEPAS 360
QY 268 TTPKEPTTTPKSAETTPKEPAPTTTKSAPTTPKSPAPTTTKPEPAPTTTKPEPAPTTTKEP 327
DB 361 TTPKEPTTTPKSAETTPKEPAPTTTKSAPTTPKSPAPTTTKPEPAPTTTKPEPAPTTTKEP 420
QY 328 APTTTTSAPTTPKEPAPTTPKPAPTTPKPEPAPTTPKPTTPPKPEPAPTTTKEPAPTTPK 387
DB 421 APTTTTSAPTTPKEPAPTTPKPAPTTPKPEPAPTTPKPTTPPKPEPAPTTTKEPAPTTPK 480
QY 388 EPAPTAPKKPAPTTPKPEPAPTTPKPEPAPTTKPSPTTPKPEPAPTTTKSAPTTKEPAPT 447
DB 481 EPAPTAPKKPAPTTPKPEPAPTTPKPEPAPTTKPSPTTPKPEPAPTTTKSAPTTKEPAPT 540
QY 448 TTKSAPTTPKPSPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTTKEPAPTTTKEP 507
DB 541 TTKSAPTTPKPSPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKPEPAPTTTKEPAPTTTKEP 600
QY 508 APTAPKEPAPTTPKETAPTTPKKLTTTPEKLAPTTPEKAPTTTPELAPTTPEEPTPTT 567
DB 601 APTAPKEPAPTTPKETAPTTPKKLTTTPEKLAPTTPEKAPTTTPELAPTTPEEPTPTT 660
QY 568 PEEPAPTTKAAAPNTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKETATTTKGTAPTTTKEP 627
DB 661 PEEPAPTTKAAAPNTPKPEPAPTTPKPEPAPTTPKPEPAPTTPKETATTTKGTAPTTTKEP 720
QY 628 APTTPKKAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKPEPAPTTPKPEPAPTTPKG 687
DB 721 APTTPKKAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKPEPAPTTPKPEPAPTTPKG 780
QY 688 TAPTTTKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKPEPAPTTPKPEPAPTTPK 747
DB 781 TAPTTTKEPAPTTPKKAPKELAPTTTKEPTSTSDKAPTTPKGTAPTTPKPEPAPTTPKPEPAPTTPK 840
QY 748 KPAPTTPEPTTTPPTSEVSTPTTTPKPEPTTIHKSDESTPELSAETTPKALENSPKPECVPT 807
DB 841 KPAPTTPEPTTTPPTSEVSTPTTTPKPEPTTIHKSDESTPELSAETTPKALENSPKPECVPT 900
QY 808 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAKPKMTETATTTESKITATTTQV 867
DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAKPKMTETATTTESKITATTTQV 960
QY 868 TSITTTQDTTTPFKITTLTKTTLAPKVTTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 927
DB 961 TSITTTQDTTTPFKITTLTKTTLAPKVTTKTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 928 PQKPTKAPKPTSTTKKPTMPRVKPKTTTTPRKMSTMPBELNPTSRIAEAMLQTTTRN 987
DB 1021 PQKPTKAPKPTSTTKKPTMPRVKPKTTTTPRKMSTMPBELNPTSRIAEAMLQTTTRN 1080
QY 988 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVPNQGIILNPMLS 1047
DB 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVPNQGIILNPMLS 1140
QY 1048 DETNLCNGKPDVGLTTLNGLTVAFRGHYFWMLSFFSPSPSPARRITEVWGIPTSIDTFT 1107
DB 1141 DETNLCNGKPDVGLTTLNGLTVAFRGHYFWMLSFFSPSPSPARRITEVWGIPTSIDTFT 1200

Query Match 96.4%; Score 6811.3; DB 13; Length 1314;
Best Local Similarity 96.0%; Pred. No. 4e-157;
Matches 1278; Conservative 4; Mismatches 12; Indels 37; Gaps 3;

QY 1 MAWKTLPIYLLILLUSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
Db |||||
QY 1 MAWKTLPIYLLILLUSVFIQVSSQDLSSCAGRCGEGYSRDATCNCYNCOHYMECCPDF 60
Db |||||
QY 61 KRVCITAEUSC-----KRCFESFERGECDCDCAQCKYDKCCPDY 100
Db |||||
QY 61 KRVCITAEUSC-----KRCFESFERGECDCDCAQCKYDKCCPDY 100
Db |||||
QY 101 ESFCAEVKDNKKNRTKKKPTKPPVWDEAGSLONGDFKVTTPDTSITTOHNVSTSPKIT 160
Db |||||
QY 108 ----LKVKDNKKNRTKKKPTKPPVWDEAGSLONGDFKVTTPDTSITTOHNVSTSPKIT 163
Db |||||
QY 161 TAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTITNNKQSTSDGKEKTTSAKETQSI 220
Db |||||
QY 164 TAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTITNNKQSTSDGKEKTTSAKETQSI 223
Db |||||
QY 221 EKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKBPTTPPKBAPTTPKBPPTTKS 280
Db |||||
QY 224 EKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKBPTTPPKBAPTTPKBPPTTKS 283
Db |||||
QY 281 APTTPKEPAPTTPKSAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 340
Db |||||
QY 284 APTTPKEPAPTTPKSAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 343
Db |||||
QY 341 EPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 400
Db |||||
QY 344 EPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 403
Db |||||
QY 401 TPKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 460
Db |||||
QY 404 TPKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 463
Db |||||
QY 461 PTTTKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 520
Db |||||
QY 464 PTTTKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 523
Db |||||
QY 521 KETAPTTPKPLPTTPPEKLAAPTTPKEPAPTTPBELAPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTP 580
Db |||||
QY 524 KETAPTTPKPLPTTPPEKLAAPTTPKEPAPTTPBELAPTTPPEEPTTPPEEPTTPPEEPTTPPEEPTTP 583
Db |||||
QY 581 PNTPEPAPTTPKEPAPTTPKBPAPTTPKETAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTP 640
Db |||||
QY 584 PNTPEPAPTTPKEPAPTTPKBPAPTTPKETAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTP 643
Db |||||
QY 641 APTTTKEPTSTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 700
Db |||||
QY 644 APTTTKEPTSTSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 703
Db |||||
QY 701 PKKPAPKELAPTTPKGTSTSDKPAPTTPKETAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 760
Db |||||
QY 704 PKKPAPKELAPTTPKGTSTSDKPAPTTPKETAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 763
Db |||||
QY 761 TSEVSTPTTPKPTTHKSPDESTPELSAEPKALENSPKBPBGVPTTKBPATKPEMTT 820
Db |||||
QY 764 TSEVSTPTTPKPTTHKSPDESTPELSAEPKALENSPKBPBGVPTTKBPATKPEMTT 823
Db |||||
QY 821 TAKDKTTERDLRTPTTPTTAAAPKMTKETAATTEKTTESKITAATTTQVSTTTQDTPPKI 880
Db |||||
QY 824 TAKDKTTERDLRTPTTPTTAAAPKMTKETAATTEKTTESKITAATTTQVSTTTQDTPPKI 883
Db |||||
QY 881 TTLKTTTLAPKVTTKKIITTEIMNKPEETAAPKDRATNSKATTPKPKQKTPKPKKPTS 940
Db |||||
QY 884 TTLKTTTLAPKVTTKKIITTEIMNKPEETAAPKDRATNSKATTPKPKQKTPKPKKPTS 943
Db |||||
QY 941 TKPKPTMPRVKPKTTPPKMTSTMPENLPTSRJAEAMLQTTTPNQTTPNSKLVENPK 1000
Db |||||
QY 944 TKPKPTMPRVKPKTTPPKMTSTMPENLPTSRJAEAMLQTTTPNQTTPNSKLVENPK 1003
Db |||||
QY 1001 SEDAGGAEGETPHMLLRPHVFWPEVTPDMDYLPRVPGIILINPMSLDETINCNGKPVGD 1060
Db |||||
QY 1004 SEDAGGAEGETPHMLLRPHVFWPEVTPDMDYLPRVPGIILINPMSLDETINCNGKPVGD 1063
Db |||||

QY 1061 LTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFTRCNCBCKTFFFKD 1120
Db |||||
QY 1064 LTTLRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPIDTVFTRCNCBCKTFFFKD 1123
Db |||||
QY 1121 SOYWRFTNDIKDAGVPRKPIFKGGLTGOIIVAALSTAKYKNWPESVYFFKGGSTQOYIY 1180
Db |||||
QY 1124 SOYWRFTNDIKDAGVPRKPIFKGGLTGOIIVAALSTAKYKNWPESVYFFKGGSTQOYIY 1183
Db |||||
QY 1181 KOEPVQKCPGRRPALNPVYVYEMTQVRRRRPERAIGPSQTHIRIOYSPARLAYODKGV 1240
Db |||||
QY 1184 KOEPVQKCPGRRPALNPVYVYEMTQVRRRRPERAIGPSQTHIRIOYSPARLAYODKGV 1243
Db |||||
QY 1241 HNEVKVILWRGLPNVVTSAISLNPKEKPDGYVYAFSKOQYINIDVPSRTARAITTRSG 1300
Db |||||
QY 1344 HNEVKVILWRGLPNVVTSAISLNPKEKPDGYVYAFSKOQYINIDVPSRTARAITTRSG 1303
Db |||||
QY 1301 QTLKSKWYNCP 1311
Db |||||
QY 1304 QTLKSKWYNCP 1314
Db |||||
RESULT 9
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Csert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1270 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44

1021 FMPEVTPDMDYLPRVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWML 1080
980 FMPEVTPDMDYLPRVNOGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWML 1039
1081 SPFPSPSPARRITEVWGIPSPIDVFTTRCNCCEGKTFEFKDSQYWFRTNDIKDAGYKPIF 1140
1040 SPFPSPSPARRITEVWGIPSPIDVFTTRCNCCEGKTFEFKDSQYWFRTNDIKDAGYKPIF 1099
1141 KGFGGLTQIIVAAALSTAKYKNWPESVYFFKRGSGIQQYIYKQEPVQKCFGRPALNYPVY 1200
1100 KGFGGLTQIIVAAALSTAKYKNWPESVYFFKRGSGIQQYIYKQEPVQKCFGRPALNYPVY 1159
1201 GEMTQVRRRRFERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSA 1260
1160 GEMTQVRRRRFERAIGPSQTHIRIQYSPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSA 1219
1261 ISLPIRKPDDGYDYAFSKDQYNNIDVSETRARITTRSGOTLSKVWYNCP 1311
1220 ISLPIRKPDDGYDYAFSKDQYNNIDVSETRARITTRSGOTLSKVWYNCP 1270

RESULT 10

US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No. US2002037894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth M.
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
; US-10-124-557-142

Query Match 96.3%; Score 6799.9; DB 13; Length 1270;
Best Local Similarity 96.9%; Pred. No. 7.2e-157;
Matches 1270; Conservative 0; Mismatches 0; Indels 41; Gaps 1;
1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYNQCQHYMECCPDF 60
1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
61 KRVCTAELSCKRCFESPERGECDDAOCKYDKCCPDYEFCAEVKDNKNRKTCKKPT 120
26 -----ELSCKRCFESPERGECDDAOCKYDKCCPDYEFCAEVKDNKNRKTCKKPT 79
121 KPXPVVVDEAGSLDGNDFKVTPTDSTTQHNVKSVSPKITTAKPINRPSLPPNSDTSKE 180
80 KPXPVVVDEAGSLDGNDFKVTPTDSTTQHNVKSVSPKITTAKPINRPSLPPNSDTSKE 139
181 TSLTVNKETTVETKETTITNKQTSIDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTP 240
140 TSLTVNKETTVETKETTITNKQTSIDGKEKTTSAKETQSIKTSADLAPTSKVLAKPTP 199
241 KAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTITTKSAPTTP 300
200 KAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTITTKSAPTTP 259
301 KEPAPTTTKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 360
260 KEPAPTTTKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 319
361 TTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 420
320 TTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 379
421 PSPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 480
380 PSPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 439
481 KKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 540
440 KKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 499
541 PTPPEKAPTTPPELAPTTPEEPPTTPPEEPPTTPPEEPPTTPPEEPPTTPPEEPPTTPPEEPPTTP 600
500 PTPPEKAPTTPPELAPTTPEEPPTTPPEEPPTTPPEEPPTTPPEEPPTTPPEEPPTTPPEEPPTTP 559
601 KEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 660
560 KEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 619
661 PKGTAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 720
620 PKGTAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 679
721 TSDKAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 780
680 TSDKAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTP 739
781 DESTPELSAETTPKALENSKPEPGVPTTKTTPAATKPEMTTTAKOKTTERDLRTTPETTTA 840
740 DESTPELSAETTPKALENSKPEPGVPTTKTTPAATKPEMTTTAKOKTTERDLRTTPETTTA 799
841 APKMTKETATTTKETSITATTTQVSTTTQDTTPPKITLTKTTLAPKVTTTKKTIIT 900
800 APKMTKETATTTKETSITATTTQVSTTTQDTTPPKITLTKTTLAPKVTTTKKTIIT 859
901 TTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPMPVRVKPKTTPTPR 960
860 TTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKPKTTPMPVRVKPKTTPTPR 919
961 KWTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSBDAGAGETPHMLLRPHV 1020
920 KWTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSBDAGAGETPHMLLRPHV 979

Query Match 96.1%; Score 6785.6; DB 13; Length 1313;
 Best Local Similarity 93.8%; Pred. No. 1.7e-156;
 Matches 1270; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWTLPTLYLLILLLSVFVIQVSSODLSSCAGRGEGYSRDATCNDYNCQHYMECCPDF 60
 DB 1 MAWTLPTLYLLILLLSVFVIQVSSQ-----25

QY 61 KRVCATLSCKGRGCFEGRGECDCDAQCKYDKCCPDYESRCAE-----106
 DB -----ELSCGRGCFEGRGECDCDAQCKYDKCCPDYESRCAEHSVSENQESSSS 79

QY 107 -----VKDNKKNRITKKPTPKPPVVDVAGSGLDNGD 137
 DB 80 SSSSSSSTIWKISSKNSAANRELQKLVKDNKKNRITKKPTPKPPVVDVAGSGLDNGD 139

QY 138 FKVTPTDSTTOHNKVSSTPKITAKPINRPSLPNSDTSKETSLLTNKETTVEKETT 197
 DB 140 FKVTPTDSTTOHNKVSSTPKITAKPINRPSLPNSDTSKETSLLTNKETTVEKETT 199

QY 198 TTNKQTSIDGKEKTSKETSQTSKTSKVLAKPTPKAETTKGPALTTTPKEP 257
 DB 200 TTNKQTSIDGKEKTSKETSQTSKTSKVLAKPTPKAETTKGPALTTTPKEP 259

QY 258 TPPTPKPASTTPKEPTPTTIKSAPTTPKBPAPTITTKSAPTTPKEPATITTKBPAPTTPK 317
 DB 260 TPPTPKPASTTPKEPTPTTIKSAPTTPKBPAPTITTKSAPTTPKEPATITTKBPAPTTPK 319

QY 318 EPAPTTPKEPATITTKSAPTTPKEPATTPKBPAPTTPKEPATTPKEPATTPKEPAT 377
 DB 320 EPAPTTPKEPATITTKSAPTTPKEPATTPKBPAPTTPKEPATTPKEPATTPKEPAT 379

QY 378 TKEPAPTTPKEPATTPKBPAPTTPKEPATTPKBPAPTTPKBPAPTTPKEPATITTKSA 437
 DB 380 TKEPAPTTPKEPATTPKBPAPTTPKEPATTPKBPAPTTPKBPAPTTPKEPATITTKSA 439

QY 438 PTTPKEPATITTKSAPTTPKEPATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 497
 DB 440 PTTPKEPATITTKSAPTTPKEPATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 499

QY 498 EPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 557
 DB 500 EPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 559

QY 558 TPTEBPTTPTEBAPTTPKAAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 617
 DB 560 TPTEBPTTPTEBAPTTPKAAAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 619

QY 618 GPAPTTLKEPATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 677
 DB 620 GPAPTTLKEPATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 679

QY 678 KEPATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 737
 DB 680 KEPATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 739

QY 738 PKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 797
 DB 740 PKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 799

QY 798 NSPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 857
 DB 800 NSPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 859

QY 858 SKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKITITTEIMNKPEETAKPKDR 917
 DB 860 SKITATTTQVSTTTQDTPPKITTLKTTTLAPKVTITTKITITTEIMNKPEETAKPKDR 919

QY 918 ATNSKATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 977
 DB 920 ATNSKATTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 979

QY 978 AMLQTTTPNQTTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVN 1037
 DB 980 AMLQTTTPNQTTPNSKLVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLRPVN 1039

QY 1038 QGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPSPARRITEVWG 1097
 DB 1040 QGIIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWMLSFPSPSPARRITEVWG 1099

QY 1098 IPSPIDTVTRCNCCEGKTFEKFQSYWRFNTNDIKDAGYKPKIFKGFGLTGQIVAAALSTA 1157
 DB 1100 IPSPIDTVTRCNCCEGKTFEKFQSYWRFNTNDIKDAGYKPKIFKGFGLTGQIVAAALSTA 1159

QY 1158 KYKNWPSVVFYFFKGGSIQYIYKQBPVQKCPGRRPALNPVYVYGMTOVRRRRRERAI GP 1217
 DB 1160 KYKNWPSVVFYFFKGGSIQYIYKQBPVQKCPGRRPALNPVYVYGMTOVRRRRRERAI GP 1219

QY 1218 SQHTIIRIQYSPARLAYODKGVHLNPKVKSILWGLPNNVTSALSLNPKPDQDYIYAF 1277
 DB 1220 SQHTIIRIQYSPARLAYODKGVHLNPKVKSILWGLPNNVTSALSLNPKPDQDYIYAF 1279

QY 1278 SKDQYNNIDVPSRTARAITTRSGQTLSKWNCP 1311
 DB 1280 SKDQYNNIDVPSRTARAITTRSGQTLSKWNCP 1313

RESULT 11
 US-10-124-557-46
 ; Sequence 46, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine
 ; Clark, Stephen C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cserr, Luann
 ; REGISTRATION NUMBER: 31,822
 ; REFERENCE/DOCKET NUMBER: GI 5190
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 876-1170
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 46:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1320 amino acids
 ; TYPE: amino acid

INFORMATION FOR SEQ ID NO: 60:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1320 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 60:
 US-10-124-557-60

Query Match 96.0%; Score 6784.9; DB 13; Length 1320;
 Best Local Similarity 93.3%; Pred. No. 1.8e-156;
 Matches 1270; Conservative 0; Mismatches 0; Indels 91; Gaps 2;
 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCVNCQHYMECCPDF 60
 1 MAWKTLPIYLLLSVFVIQVSSQ-----25
 61 KRVCATLSKGRCFESFERGECDCDAQCKYDKCCPDYESFCA-----105
 26 -----ELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
 106 -----EYKDNKNRTKKKPTKPPVVDDEAG 130
 80 PPSGASQTIKTSRSPKPNKKTKVIESEBITEVKDNKNRTKKKPTKPPVVDDEAG 139
 131 SGLDNGDFKVTDTSTTQHNKSVSTSPKITTAKPINRPSLPNNSDTSKETSITVNKETT 190
 140 SGLDNGDFKVTDTSTTQHNKSVSTSPKITTAKPINRPSLPNNSDTSKETSITVNKETT 199
 191 VETKETITNNKQISTDGKEXTSAKETQSIKTSKADLAPTSKVLAKEPTTKEGPA 250
 200 VETKETITNNKQISTDGKEXTSAKETQSIKTSKADLAPTSKVLAKEPTTKEGPA 259
 251 LTPKPEPTTTPKEPASTTTPKEPTTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 310
 260 LTPKPEPTTTPKEPASTTTPKEPTTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 319
 311 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 370
 320 PAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 379
 371 PKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 430
 380 PKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 439
 431 PTTTSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 490
 440 PTTTSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 499
 491 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 550
 500 PAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 559
 551 TPEELAPTTPEEPTTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 610
 560 TPEELAPTTPEEPTTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 619
 611 TAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 670
 620 TAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE 679
 671 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 730
 680 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 739
 731 KETAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 790
 740 KETAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKE 799
 791 PTPKALENSPKPGVPTTKTAAKPEMTTAKDKTERDITTTETTTAAAPKMTKETAT 850
 800 PTPKALENSPKPGVPTTKTAAKPEMTTAKDKTERDITTTETTTAAAPKMTKETAT 859

851 TTEKTESKITATTTQVTSITTTQDTPPKITLTLTKTTLAPKVTTTKITTTIEMNKPEE 910
 860 TTEKTESKITATTTQVTSITTTQDTPPKITLTLTKTTLAPKVTTTKITTTIEMNKPEE 919
 911 TAKPKDRATNSKATTPKPKETKAPKPTSTKSKTMRVRKPKTTPPRKWTSTMPBLN 970
 920 TAKPKDRATNSKATTPKPKETKAPKPTSTKSKTMRVRKPKTTPPRKWTSTMPBLN 979
 971 PTSRIAEAMLQTTTRPNQTPNSKLVEVNPXKSEDAGAGETPHMLLRPHVEMPEVTPMD 1030
 980 PTSRIAEAMLQTTTRPNQTPNSKLVEVNPXKSEDAGAGETPHMLLRPHVEMPEVTPMD 1039
 1031 YLPRVFNQGIILNMLSDENICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1090
 1040 YLPRVFNQGIILNMLSDENICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPFPSPSPAR 1099
 1091 RITEVWGIIPSPIDVTRCNCEGKTFEFGGSIQYIYKQEPVQKCPGRPALNYPVYGEVTVRRR 1150
 1100 RITEVWGIIPSPIDVTRCNCEGKTFEFGGSIQYIYKQEPVQKCPGRPALNYPVYGEVTVRRR 1159
 1151 VAALSTAKYKNWPESVYFFKRGSIQYIYKQEPVQKCPGRPALNYPVYGEVTVRRR 1210
 1160 VAALSTAKYKNWPESVYFFKRGSIQYIYKQEPVQKCPGRPALNYPVYGEVTVRRR 1219
 1211 FERAIGPSQTHIRIQYSPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPIRKP 1270
 1220 FERAIGPSQTHIRIQYSPARLAYQDKGVHNEVKVSIILWRGLPNVVTSAISLPIRKP 1279
 1271 GYDYAFSKDQYINIDVPSRTARITTRSGQTLISKWYNCP 1311
 1280 GYDYAFSKDQYINIDVPSRTARITTRSGQTLISKWYNCP 1320

RESULT 13
 US-10-124-557-52
 ; Sequence 52, Application US/10124557
 ; Publication No. US20020137894A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Turner, Katherine C.
 ; Jacobs, Kenneth
 ; Hewick, Rodney M.
 ; Gesner, Thomas G.
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
 ; NUMBER OF SEQUENCES: 143
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/124,557
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/643,502
 ; FILING DATE: 18-JAN-1991
 ; APPLICATION NUMBER: US 07/546,114
 ; FILING DATE: 29-JUN-1990
 ; APPLICATION NUMBER: US 07/457,196
 ; FILING DATE: 29-DEC-1989
 ; APPLICATION NUMBER: US 07/390,901
 ; FILING DATE: 08-AUG-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cseert, Luann
 ; REGISTRATION NUMBER: 31,822

| | | | |
|---|------|--|------|
| REFERENCE/DOCKET NUMBER: GI 5190 | | | |
| TELECOMMUNICATION INFORMATION: | | | |
| TELEPHONE: (617) 876-1170 | | | |
| TELEFAX: (617) 876-5851 | | | |
| INFORMATION FOR SEQ ID NO: 52: | | | |
| SEQUENCE CHARACTERISTICS: | | | |
| LENGTH: 1363 amino acids | | | |
| TYPE: amino acid | | | |
| TOPOLOGY: linear | | | |
| MOLECULE TYPE: protein | | | |
| SEQUENCE DESCRIPTION: SEQ ID NO: 52: | | | |
| US-10-124-557-52 | | | |
| Query Match | | | |
| Best Local Similarity 96.0%; Score 6780.6; DB 13; Length 1363; | | | |
| Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2; | | | |
| QY | 1 | MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSDATCNCNDYNCQHYMECCPDF | 60 |
| DB | 1 | MAWKTLPIYLLLLSVFVIQQVSSQ----- | 25 |
| QY | 61 | KRVCTAELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- | 106 |
| DB | 26 | -----ELSCGRCFESFERGECDCDAQCKKYDKCCPDYESFCAE----- | 79 |
| QY | 107 | ----- | 106 |
| DB | 80 | PPSGASQTIKSTTKRSPKPPNKKTKVIESEBITEHSVSENQESSSSSSSSSTIW | 139 |
| QY | 107 | -----VKDNKQRTKKKPKPPVVDVDEAGSLDNGPKVTTTDTST | 147 |
| DB | 140 | KIKSSKNSAANRELQKKLVKDNKKNKTKKKPKPPVVDVDEAGSLDNGPKVTTTDTST | 199 |
| QY | 148 | TOHNKYVSTPKIITAKPINRPSLPNSDTSKETSIVNKKETTIVETKTTTNKQSTDG | 207 |
| DB | 200 | TOHNKYVSTPKIITAKPINRPSLPNSDTSKETSIVNKKETTIVETKTTTNKQSTDG | 259 |
| QY | 208 | KEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS | 267 |
| DB | 260 | KEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTTTPKEPAS | 319 |
| QY | 268 | TTKEPTTTPKSAPTTPKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEP | 327 |
| DB | 320 | TTKEPTTTPKSAPTTPKEPATTTTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEP | 379 |
| QY | 328 | APTTPKSAPTTTKEPATTTPKKPAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTPK | 387 |
| DB | 380 | APTTPKSAPTTTKEPATTTPKKPAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTPK | 439 |
| QY | 388 | EPAPTAPKAPAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT | 447 |
| DB | 440 | EPAPTAPKAPAPTTPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT | 499 |
| QY | 448 | TTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEP | 507 |
| DB | 500 | TTKSAPTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEP | 559 |
| QY | 508 | APTAPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT | 567 |
| DB | 560 | APTAPKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTT | 619 |
| QY | 568 | PEEPAPTPKAAANTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT | 627 |
| DB | 620 | PEEPAPTPKAAANTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPAT | 679 |
| QY | 628 | APTTPKAPKELAPTTTKEPATTTTSDKAPTTTKEPATTTTKEPATTTTKEPATTTTKEPAT | 687 |
| DB | 680 | APTTPKAPKELAPTTTKEPATTTTSDKAPTTTKEPATTTTKEPATTTTKEPATTTTKEPAT | 739 |
| QY | 688 | TAPTTLKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTK | 747 |
| DB | 740 | TAPTTLKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTTKEPATTTK | 799 |
| QY | 748 | KPAPTTPEPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPPTKALENSPKPEGVPT | 807 |
| DB | 800 | KPAPTTPEPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPPTKALENSPKPEGVPT | 859 |
| QY | 808 | TKTAAATKPEMTTAKDKTTERDLATTTTAAAPKMTKETATTTTETTESKITATTTOV | 867 |
| DB | 860 | TKTAAATKPEMTTAKDKTTERDLATTTTAAAPKMTKETATTTTETTESKITATTTOV | 919 |
| QY | 868 | TSSTTTQDTPPKITTLTKTLAPKVTTLTKTLTTEIMNKPEETAKPKDRATNSKATTPK | 927 |
| DB | 920 | TSSTTTQDTPPKITTLTKTLAPKVTTLTKTLTTEIMNKPEETAKPKDRATNSKATTPK | 979 |
| QY | 928 | POKPKAPKPKPTSTKKPKTMRVRPKTTPTPRKMSTIMPELNPTSRISAEAMLQTTTRN | 987 |
| DB | 980 | POKPKAPKPKPTSTKKPKTMRVRPKTTPTPRKMSTIMPELNPTSRISAEAMLQTTTRN | 1039 |
| QY | 988 | QTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMDYLRVNPQGGIINPMLS | 1047 |
| DB | 1040 | QTPNSKLVNPKSDEAGAGETPHMLLRPHVFMPEVTPDMDYLRVNPQGGIINPMLS | 1099 |
| QY | 1048 | DETNICNGKPVDLGTLTLRNGTLVAPRGHYFWMLSPSPSPARRITEVWGIPSPIDTFT | 1107 |
| DB | 1100 | DETNICNGKPVDLGTLTLRNGTLVAPRGHYFWMLSPSPSPARRITEVWGIPSPIDTFT | 1159 |
| QY | 1108 | RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIVAAALSTAKYKNWPESVY | 1167 |
| DB | 1160 | RCNCEGKTFFFKDSQYWRFTNDIKDAGYKPIPKFGGLTGQIVAAALSTAKYKNWPESVY | 1219 |
| QY | 1168 | FFKRGSGIOQYIKQBPVQKCGRRPALNYPVYVGTQVRRRFFERAIGPSQTHTRIQY | 1227 |
| DB | 1220 | FFKRGSGIOQYIKQBPVQKCGRRPALNYPVYVGTQVRRRFFERAIGPSQTHTRIQY | 1279 |
| QY | 1228 | SPARLAYQDKVGLHNEKVSILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYYNIDV | 1287 |
| DB | 1280 | SPARLAYQDKVGLHNEKVSILWRGLPNVVTSAISLPNIRKPDGYDYAFSKDQYYNIDV | 1339 |
| QY | 1288 | PSRTARAITRSQGLTSKVWYNCP 1311 | |
| DB | 1340 | PSRTARAITRSQGLTSKVWYNCP 1363 | |
| RESULT 14 | | | |
| US-10-124-557-104 | | | |
| ; Sequence 104, Application US/10124557 | | | |
| ; Publication No. US20020137894A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Turner, Katherine | | | |
| ; Clark, Stephen C. | | | |
| ; Jacobs, Kenneth | | | |
| ; Hewick, Rodney M. | | | |
| ; Gesner, Thomas G. | | | |
| ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors | | | |
| ; NUMBER OF SEQUENCES: 143 | | | |
| ; CORRESPONDENCE ADDRESS: | | | |
| ; ADDRESSEE: Genetics Institute, Inc. | | | |
| ; STREET: 87 CambridgePark Drive | | | |
| ; CITY: Cambridge | | | |
| ; STATE: Massachusetts | | | |
| ; COUNTRY: U.S.A. | | | |
| ; ZIP: 02140 | | | |
| ; COMPUTER READABLE FORM: | | | |
| ; MEDIUM TYPE: Floppy disk | | | |
| ; COMPUTER: IBM PC compatible | | | |
| ; OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| ; SOFTWARE: PatentIn Release #1.0, Version #1.25 | | | |
| ; CURRENT APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US/10/124,557 | | | |
| ; FILING DATE: 16-Apr-2002 | | | |
| ; CLASSIFICATION: <Unknown> | | | |
| ; PRIOR APPLICATION DATA: | | | |
| ; APPLICATION NUMBER: US 07/643,502 | | | |
| ; FILING DATE: 18-JAN-1991 | | | |
| ; APPLICATION NUMBER: US 07/546,114 | | | |

```

; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match          79.3%; Score 5603.7; DB 13; Length 1140;
Best Local Similarity 91.8%; Pred No. 6.4e-128;
Matches 1047; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
DB 1 MAWKTLPIYLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY 61 KRVTAEISCKGRCFESFREGRECDCAQCKYDKCCPDYESFCAE-----106
DB 61 KRVTAEISCKGRCFESFREGRECDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY 107 -----106
DB 121 PPSGASQTIKSTKRSKPPNKKTKKVISEBITEHSVSENQESSSSSSSSSTIW 180
QY 107 -----VKNNKNTTKKPPKPPVVDVDEAGSLDNGDFKVTTPDTS 147
DB 181 KIKSSKNSAANRELQKKLVKONKKNRTKKKPTPKPPVVDVDEAGSLDNGDFKVTTPDTS 240
QY 148 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTINKQSTDG 207
DB 241 TQHNKVTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTINKQSTDG 300
QY 208 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKBPPTTPKBPAS 267
DB 301 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKBPPTTPKBPAS 360
QY 268 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEP 327
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEP 420
QY 328 APTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAP 387
DB 421 APTTTKSAPTTPKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAP 480
QY 388 EPAPTAPKKPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKSAPTTTKEPAPT 447
DB 481 EPAPTAPKKPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKSAPTTTKEPAPT 540
QY 448 TTKSAPTTPKSPPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKP 507
DB 541 TTKSAPTTPKSPPTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKP 600
QY 508 APTAPKEPAPTTKETAATPKLTPTEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 567
DB 601 APTAPKEPAPTTKETAATPKLTPTEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 660
QY 568 PEEAPPTPKAAAPNTPKBPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 627
DB 661 PEEAPPTPKAAAPNTPKBPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEP 720

```

```

QY 628 APTTPKPKAPKELAPTTTKKEPTSTTSKAPAPTTTPKGTAPTTPKBPAPTTPKBP 687
DB 721 APTTPKPKAPKELAPTTTKKEPTSTTSKAPAPTTTPKGTAPTTPKBPAPTTPKBP 780
QY 688 TAPTTLKEPAPTTTPKBPAPKELAPTTTKGPTSTTSKAPAPTTTPKBPAPTTPK 747
DB 781 TAPTTLKEPAPTTTPKBPAPKELAPTTTKGPTSTTSKAPAPTTTPKBPAPTTPK 840
QY 748 KPAPTTTPETPPPTTSEVSTPTTKKEPTTIHKSPDESTPELSAETPKALENSPK 807
DB 841 KPAPTTTPETPPPTTSEVSTPTTKKEPTTIHKSPDESTPELSAETPKALENSPK 900
QY 808 TKTPAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETATTTKTESKITATT 867
DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETATTTKTESKITATT 960
QY 868 TSTTTQDTTPFKITTLAKTITLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSK 927
DB 961 TSTTTQDTTPFKITTLAKTITLAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSK 1020
QY 928 POKPTKAPKKPTSTKPKTMPRVKPKTTTPRKMSTSTMPELNPTSRIAEAMLOTT 987
DB 1021 POKPTKAPKKPTSTKPKTMPRVKPKTTTPRKMSTSTMPELNPTSRIAEAMLOTT 1080
QY 988 QTPNSKLIVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMMDYLPVFNQGIINP 1047
DB 1081 QTPNSKLIVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMMDYLPVFNQGIINP 1140

```

RESULT 15

```

US-10-124-557-74
; Sequence 74, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851

```

```

; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1038 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match          77.9%; Score 5500.7; DB 13; Length 1038;
Best Local Similarity 79.2%; Pred. No. 1.8e-125;
Matches 1038; Conservative 0; Mismatches 0; Indels 273; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGGYGRDATCNDYNCQHMECCPDF 60
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ-----25
QY 61 KRVCTAELCKGRCHESFERGRECDCAQCKYDKCCPDYESFCAEVKDNKNRTKKKPT 120
Db -----ELCKGRCHESFERGRECDCAQCKYDKCCPDYESFCAEVKDNKNRTKKKPT 79
QY 121 KPVPVDEAGSLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPPNSDTSKE 180
Db 80 KPVPVDEAGSLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINRPSLPPNSDTSKE 139
QY 181 TSLTVNKETTVETKETTTNNKOTSDGKEKTSKAKETQSIKTSKADLAPTSKVLAKPTP 240
Db 140 TSLTVNKETTVETKETTTNNKOTSDGKEKTSKAKETQSIKTSKADLAPTSKVLAKPTP 199
QY 241 KAEITTKGPAITTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTP 300
Db 200 KAEITTKGPAITTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTP 259
QY 301 KEPAITTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAP 360
Db 260 KEPAITTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAP 319
QY 361 TTPKEPTTPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 420
Db 320 TTPKEPTTPTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKE 379
QY 421 PSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTP 480
Db 380 PSPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTP 439
QY 481 KKPAITTPKEPAPTTTPKEPAPTTTKPAAPAPKEPAPTTPKETAPTTPKKLTPTTPKLA 540
Db 440 KKPAITTPKEPAPTTTPKEPAPTTTKPAAPAPKEPAPTTPKETAPTTPKKLTPTTPKLA 499
QY 541 PTTPEKAPAPTTPEELAPTTPEEPTTPPEEAPPTTPKAAAPNTTPKEPAPTTTPKEPAP 600
Db 500 PTTPEKAPAPTTPEELAPTTPEEPTTPPEEAPPTTPKAAAPNTTPKEPAPTTTPKEPAP 559
QY 601 KEPAITTPKEAPTTTPKGTAPTTLLKEPAPTTPKPAKELAPTTTKPTSTTSKAPATT 660
Db 560 KEPAITTPKEAPTTTPKGTAPTTLLKEPAPTTPKPAKELAPTTTKPTSTTSKAPATT 619
QY 661 PKGTAPTTPKPAKELAPTTPKGTAPTTLLKEPAPTTTPKKAPKELAPTTTKGPTST 720
Db 620 PKGTAPTTPKPAKELAPTTPKGTAPTTLLKEPAPTTTPKKAPKELAPTTTKGPTST 679
QY 721 TSDKAPATTTPKEAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTIHKSP 780
Db 680 TSDKAPATTTPKEAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTIHKSP 739
QY 781 DESTPELSAEPPTPKALENSKPEPGVPTTKTAAATKPEMTTTAKOKITTERDLRTTPETTTA 840
Db 740 DESTPELSAEPPTPKALENSKPEPGVPTTKTAAATKPEMTTTAKOKITTERDLRTTPETTTA 799
QY 841 APKMTKETATTTTEKTESKITATTTQVTSSTTQDTPPKITTLKTTTAPKVTITTKKIT 900
Db 800 APKMTKETATTTTEKTESKITATTTQVTSSTTQDTPPKITTLKTTTAPKVTITTKKIT 859
```

Search completed: October 13, 2004, 11:53:32
Job time : 142.379 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 144.144 Seconds
(without alignments)
5233.063 Million cell updates/sec

Title: SEQ1-H
Perfect score: 7064
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....ARAITRRSGTSLSKVWVNCPL 1311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1825181 segs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_02:*
1: uniprot_prot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|------------|----------------------|
| 1 | 7044.7 | 99.7 | 1404 | 2 | Q92954 | Q92954 homo sapien |
| 2 | 7029.7 | 99.5 | 1404 | 2 | Q9BX49 | Q9BX49 homo sapien |
| 3 | 4370.6 | 61.9 | 933 | 2 | Q6ZM25 | Q6ZM25 homo sapien |
| 4 | 4370.6 | 61.9 | 933 | 2 | BAD18580 | BAD18580 h cdna fl |
| 5 | 3827.7 | 54.2 | 1054 | 2 | Q9JM99 | Q9JM99 mus musculus |
| 6 | 1714.7 | 24.3 | 401 | 2 | O77765 | O77765 bos taurus |
| 7 | 1398.2 | 19.8 | 5179 | 1 | MUC2_HUMAN | MUC2_HUMAN |
| 8 | 1338 | 18.9 | 1225 | 2 | Q9VR49 | Q9VR49 |
| 9 | 1333 | 18.9 | 251 | 2 | Q6DNC4 | Q6DNC4 |
| 10 | 1311.6 | 18.6 | 1761 | 2 | AA564673 | AA564673 drosophila |
| 11 | 1311.6 | 18.6 | 1761 | 2 | Q7KTF6 | Q7KTF6 |
| 12 | 1271.2 | 18.0 | 3150 | 2 | Q7PMD5 | Q7PMD5 |
| 13 | 1211.6 | 17.2 | 3409 | 2 | Q6SSE6 | Q6SSE6 chlamydomon |
| 14 | 1211.6 | 17.2 | 3409 | 2 | AA507044 | AA507044 chlamydomon |
| 15 | 1183 | 16.7 | 251 | 2 | Q6QJF3 | Q6QJF3 bos taurus |
| 16 | 1183 | 16.7 | 251 | 2 | AA583977 | AA583977 bos taurus |
| 17 | 1154.2 | 16.3 | 1664 | 1 | SUP1_CLOTH | SUP1_CLOTH |
| 18 | 1150.6 | 16.3 | 3889 | 2 | Q6SSE8 | Q6SSE8 clostridium |
| 19 | 1150.6 | 16.3 | 3889 | 2 | AA507042 | AA507042 chlamydomon |
| 20 | 1112.2 | 15.7 | 5703 | 1 | MUSB_HUMAN | MUSB_HUMAN |
| 21 | 1102.5 | 15.6 | 1349 | 2 | Q8WQ4 | Q8WQ4 homo sapien |
| 22 | 1102 | 15.6 | 3432 | 2 | Q8TR51 | Q8TR51 drosophila |
| 23 | 1102 | 15.6 | 3458 | 2 | Q8TR52 | Q8TR52 drosophila |
| 24 | 1096.4 | 15.5 | 1795 | 2 | Q6894 | Q6894 drosophila |
| 25 | 1093.3 | 15.5 | 1079 | 2 | Q9N4S7 | Q9N4S7 caenorhabdi |
| 26 | 1084.6 | 15.4 | 9234 | 2 | Q7KTP5 | Q7KTP5 |
| 27 | 1084.6 | 15.4 | 9234 | 2 | AA10531 | AA10531 drosophila |
| 28 | 1079 | 15.3 | 23015 | 2 | Q8IQ18 | Q8IQ18 drosophila |
| 29 | 1079 | 15.3 | 23015 | 2 | AA10358 | AA10358 drosophila |
| 30 | 1061.7 | 15.0 | 34350 | 2 | Q8WZ42 | Q8WZ42 homo sapien |
| 31 | 1054.5 | 14.9 | 2284 | 2 | Q9VPG1 | Q9VPG1 drosophila |

32 1038 14.7 7962 2 Q10465
33 1034.5 14.6 972 2 Q7QKK7
34 1034.2 14.6 1607 2 Q8H6Q5
35 1034.2 14.6 1607 2 AAP74661
36 1021.4 14.5 2187 2 P70670
37 1011.3 14.3 1489 2 O96449
38 1010.5 14.3 1274 2 Q20007
39 1001.8 14.2 1458 2 Q757N5
40 1001.8 14.2 1458 2 AAS52662
41 1001.8 14.2 4498 2 Q9W223
42 983.6 13.9 10578 2 Q8ISF5
43 983.2 13.9 1480 2 Q9LIE8
44 981.6 13.9 18519 2 Q8ISF6
45 981.6 13.9 18534 2 Q8ISF7

ALIGNMENTS

RESULT 1

Q92954 PRELIMINARY; PRT; 1404 AA.
AC Q92954;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Megakaryocyte stimulating factor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
RT "Purification, Biochemical Characterization, and Cloning of a Novel
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony
RT Stimulating Activity.";
RL Blood 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,
RA Jacobs K., Turner K.;
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL Mosher D.F. (eds.);
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier
RL Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; U70136; AAB09089.1; -.
DR HSSP; P04004; 10C0.
DR Genew; HGNC:9364; PRG4.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin B; 2.
DR PRINTS; PS00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

| | | | | | | |
|-----------------------|-----|--|---------------------|------------|--------------|--------------------|
| Query Match | | 99.7%; | Score 7044.7; | DB 2; | Length 1404; | |
| Best Local Similarity | | 93.4%; | Pred. No. 1.4e-139; | | | |
| Matches 1311; | | Conservative | 0; | Mismatches | 0; | Indels 93; Gaps 1; |
| Qy | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSDATCNCYNCHYMCCPDF | 60 | | | |
| Db | 1 | MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSDATCNCYNCHYMCCPDF | 60 | | | |
| Qy | 61 | KEVCTAELSCKRCFESPERGECDCDAQCKYDKCCPDYESFCAE----- | 106 | | | |
| Db | 61 | KEVCTAELSCKRCFESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP | 120 | | | |
| Qy | 107 | ----- | 106 | | | |
| Db | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSTIW | 180 | | | |
| Qy | 107 | -----VKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTS | 147 | | | |
| Db | 181 | KIKSSKNSAANRELQKKLVKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTS | 240 | | | |
| Qy | 148 | TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETKETTINKQTS | 207 | | | |
| Db | 241 | TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTETKETTINKQTS | 300 | | | |
| Qy | 208 | KEKTSKETSQSTKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKEBAS | 267 | | | |
| Db | 301 | KEKTSKETSQSTKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKEBAS | 360 | | | |
| Qy | 268 | TTPKETPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTPKE | 327 | | | |
| Db | 361 | TTPKETPTTIKSAPTTTPKEPATTTTKSAPTTTPKEPATTTTPKEPATTTTPKE | 420 | | | |
| Qy | 328 | APTPTKSAPTTTPKEPATTTPKKPAPTTTPKEPATTTTPKEPATTTTPKEPATTTPK | 387 | | | |
| Db | 421 | APTPTKSAPTTTPKEPATTTPKKPAPTTTPKEPATTTTPKEPATTTTPKEPATTTPK | 480 | | | |
| Qy | 388 | EPAPTAPKKAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT | 447 | | | |
| Db | 481 | EPAPTAPKKAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT | 540 | | | |
| Qy | 448 | TTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKKP | 507 | | | |
| Db | 541 | TTKSAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKKP | 600 | | | |
| Qy | 508 | APTAPKEPATTTTPKETAPTTPKLTPTTPEKLAPTTTPKEPATTTTPKEPATTTTPEEPT | 567 | | | |
| Db | 601 | APTAPKEPATTTTPKETAPTTPKLTPTTPEKLAPTTTPKEPATTTTPKEPATTTTPEEPT | 660 | | | |
| Qy | 568 | PEEPAPTTPKAAANTPKAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKEP | 627 | | | |
| Db | 661 | PEEPAPTTPKAAANTPKAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTKEP | 720 | | | |
| Qy | 628 | APTTPKPAKELAPTTTPKETSTTSOKPAPTTTPKGATTTTPKEPATTTTPKEPATTTPKG | 687 | | | |
| Db | 721 | APTTPKPAKELAPTTTPKETSTTSOKPAPTTTPKGATTTTPKEPATTTTPKEPATTTPKG | 780 | | | |
| Qy | 688 | TAPTTLKEPATTTTPKKAPKELAPTTTKGPTSTTSOKPAPTTTPKETATTTTPKEPATTTPK | 747 | | | |
| Db | 781 | TAPTTLKEPATTTTPKKAPKELAPTTTKGPTSTTSOKPAPTTTPKETATTTTPKEPATTTPK | 840 | | | |
| Qy | 748 | KPATTTTTPPTTSEVSTPTTTPKETTTIHKSPDESTPELSAETPKALENSPKPEGVPT | 807 | | | |
| Db | 841 | KPATTTTTPPTTSEVSTPTTTPKETTTIHKSPDESTPELSAETPKALENSPKPEGVPT | 900 | | | |
| Qy | 808 | TKTAAATPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITATTTQV | 867 | | | |
| Db | 901 | TKTAAATPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITATTTQV | 960 | | | |
| Qy | 868 | TSTTTQDTPPKFITTLLTKTTLAPKVTTKKITTTEIMNKBEETAKPKORATNSKATTPK | 927 | | | |
| Db | 961 | TSTTTQDTPPKFITTLLTKTTLAPKVTTKKITTTEIMNKBEETAKPKORATNSKATTPK | 1020 | | | |
| Qy | 928 | FOKPTKAPKPKTSTTKKPTKMPVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRFN | 987 | | | |
| Db | 121 | PPSGASQTIKSTTKRSPKPNKKTKVIESEITEHSVSENQESSSSSSSSSTIR | 180 | | | |

QY 107 -----VKDNKNRTKKXPTPKPPVVDAGSLDNGDFKVTPTDST 147
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKXPTPKPPVVDAGSLDNGDFKVTPTDST 240
QY 148 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 207
Db 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 300
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKBPAS 267
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKBPAS 360
QY 268 TTPKEPTPTTIKSAPTTPKBPAPTTPKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 327
Db 361 TTPKEPTPTTIKSAPTTPKBPAPTTPKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 420
QY 328 APTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 387
Db 421 APTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 480
QY 388 EPAPTAPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 447
Db 481 EPAPTAPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 540
QY 448 TTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 507
Db 541 TTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPT 600
QY 508 APTAPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 567
Db 601 APTAPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 660
QY 568 PEPAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 627
Db 661 PEPAPTTPKAAAPNTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 720
QY 628 APTTPKBPAPKELAPTTTKEPTSTTSKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 687
Db 721 APTTPKBPAPKELAPTTTKEPTSTTSKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 780
QY 688 TAPTTLKEBPAPTTPKBPAPKELAPTTTKEPTSTTSKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 747
Db 781 TAPTTLKEBPAPTTPKBPAPKELAPTTTKEPTSTTSKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 840
QY 748 KPAPTTPPTTSEVSTPTTKEPTTHKSDPTTKEPTTHKSDPTTKEPTTHKSDPTTKEPTTHKSDPTT 807
Db 841 KPAPTTPPTTSEVSTPTTKEPTTHKSDPTTKEPTTHKSDPTTKEPTTHKSDPTTKEPTTHKSDPTT 900
QY 808 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTTETKTTESKITATTTOV 867
Db 901 TKTPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTTTETKTTESKITATTTOV 960
QY 868 TSTTTQDTPPKITLKITTLAPKVTITTKITTTTMMKPBETAKPKDRATNSKATTPK 927
Db 961 TSTTTQDTPPKITLKITTLAPKVTITTKITTTTMMKPBETAKPKDRATNSKATTPK 1020
QY 928 POKPTKAPKAPKPTSTKPKTPMPVRKPTTTPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMT 987
Db 1021 POKPTKAPKAPKPTSTKPKTPMPVRKPTTTPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMT 1080
QY 988 QTPNSKLVENPKSBDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1047
Db 1081 QTPNSKLVENPKSBDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1140
QY 1048 DETNTCNGKPVUGLTLTRNGTLVAFRGHYFWMLSDFSPSPARRITEVWGIPSIDTFT 1107
Db 1141 DETNTCNGKPVUGLTLTRNGTLVAFRGHYFWMLSDFSPSPARRITEVWGIPSIDTFT 1200
QY 1108 RCNCEGKTFFFKDSQVWFTNDIKDAGYKPKTFKFGGLTGQIVAAALSTAKYKNWPESVY 1167
Db 1201 RCNCEGKTFFFKDSQVWFTNDIKDAGYKPKTFKFGGLTGQIVAAALSTAKYKNWPESVY 1260

QY 1168 FFKGGSIQIYYIKQEPVQKCPGRRPALNVPVYGMTQVRRRRRFFERAIGPSQTHIRIQY 1227
Db 1261 FFKGGSIQIYYIKQEPVQKCPGRRPALNVPVYGMTQVRRRRRFFERAIGPSQTHIRIQY 1320
QY 1228 SPARLAYQDKGVLHNEVKVSTLWGLNPNVVTSAISLNRKPDGYDYAFSKDQYINIDV 1287
Db 1321 SPARLAYQDKGVLHNEVKVSTLWGLNPNVVTSAISLNRKPDGYDYAFSKDQYINIDV 1380
QY 1288 PSRTARAITTRSGQTLSSKVVNCP 1311
Db 1381 PSRTARAITTRSGQTLSSKVVNCP 1404
RESULT 3
Q6ZMZ5 PRELIMINARY; PRT; 933 AA.
AC Q6ZMZ5; 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein FLJ16561.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsubi T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Magatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK131434; BAD18580.1; --
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin_B; 1.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 1.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 1.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;
Query Match 61.9%; Score 4370.6; DB 2; Length 933;
Best Local Similarity 59.6%; Pred. No. 1e-83;
Matches 839; Conservative 0; Mismatches 1; Indels 564; Gaps 3;
QY 1 MAWKTLPIYLLLLLSVFIQVSSODLSSCAGCGGYSRDATCNCYDNCQHYMECCPDF 60
Db 1 MAWKTLPIYLLLLLSVFIQVSSO----- 25
QY 61 KRVTAEISCKGRCPESPERGECDCDAQCKYDKCCPDYESFCAE----- 106
Db 26 -----ELSCGRCPESPERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
QY 107 ----- 106
Db 80 PPSGASQTIKSTTKSPKPPNKKTKVIESEITEHSVSNQSSSSSSSSSSSTIR 139
QY 107 -----VKDNKNRTKKXPTPKPPVVDAGSLDNGDFKVTPTDST 147
Db 140 KIKSSKNSAANRELQKKLVKDNKNRTKKXPTPKPPVVDAGSLDNGDFKVTPTDST 199
QY 148 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 207
Db 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKEITVETKETTNNKQISTDG 259
QY 208 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTPPKBPAS 267

| | | | |
|----|------|--|------|
| Db | 910 | PSRTARAITTRSGQTLKVVWVNC | 933 |
| Db | 1260 | KEKTTSAKETQSIEKTSKDLAPTSKVLAKPTPKAEITTKGPAITTPKEPTTTTKEPAS | 319 |
| Qy | 268 | TTTKEPTPTTIKSAPTTKEPAPTTTTSAPTTKPKBPAPTTTKEPAPTTTKEPAPTTTKEP | 327 |
| Db | 320 | TTTKEPTPTTIKSAPTTKEPAPTTTTSAPTTKPKBPAPTTTKEPAPTTTKEPAPTTTKEP | 370 |
| Qy | 328 | APTPTTSAPTTKEPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTTTTPK | 387 |
| Db | 371 | ----- | 370 |
| Qy | 388 | EPAPTAPKBPAPTTTKEPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTT | 447 |
| Db | 371 | ----- | 370 |
| Qy | 448 | TTKSAPTTTPKPSPTTTKEPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTT | 507 |
| Db | 371 | ----- | 370 |
| Qy | 508 | APTAPKEPAPTTTPKETAPTTPKLTPTTPEKLAPTTPEKAPTTTPEELAPTTPEEPTPTT | 567 |
| Db | 371 | ----- | 370 |
| Qy | 568 | PBEAPTTPKAAPNTPKBPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTTTTPKBPAPTTTTPK | 627 |
| Db | 371 | ----- | 370 |
| Qy | 628 | APTTPKBPAPKELAPTTTKEPTSTSDKAPTTTPKGTAPTTTPKBPAPTTTTPKBPAPTTTPKG | 687 |
| Db | 371 | ----- | 370 |
| Qy | 688 | TAPTTLKEPAPTTTPKBPAPKELAPTTTKEPTSTSDKAPTTTTPKGTAPTTTPKBPAPTTTPK | 747 |
| Db | 371 | ----- | 370 |
| Qy | 748 | KPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSPKPGVPT | 807 |
| Db | 371 | -PAPTTPETPPPTTSEVSTPTTKEPTTIHKSPDESTPELSABPTPKALENSPKPGVPT | 429 |
| Qy | 808 | TKTPAATKPEMTTAKDKTTERDLRTPTTTTAAPKMTKETATTTTEKTTESKITATTQV | 867 |
| Db | 430 | TKTPAATKPEMTTAKDKTTERDLRTPTTTTAAPKMTKETATTTTEKTTESKITATTQV | 489 |
| Qy | 868 | TSITTQDTPPKTTLTKTTLAPKVTTTKTITTEIMNKPEETAKPKDRATNSKATTPK | 927 |
| Db | 490 | TSITTQDTPPKTTLTKTTLAPKVTTTKTITTEIMNKPEETAKPKDRATNSKATTPK | 549 |
| Qy | 928 | POKETKAPKPTSTPKPKTTPRVKPKTTPTPRKTSTMPELNPTSRIAEAMLQTTTRPN | 987 |
| Db | 550 | POKETKAPKPTSTPKPKTTPRVKPKTTPTPRKTSTMPELNPTSRIAEAMLQTTTRPN | 609 |
| Qy | 988 | QTPNSKLVEVNPKSEDAGGAETPHMLLRPHVFMPEVTPDMDYLPVPNOGIIINPMLS | 1047 |
| Db | 610 | QTPNSKLVEVNPKSEDAGGAETPHMLLRPHVFMPEVTPDMDYLPVPNOGIIINPMLS | 669 |
| Qy | 1048 | DETNI CNKGPVDGLTTLNGLTVAFRGHYFWMLSFPSPSPSAPRITTEVWGTPSIDIVFT | 1107 |
| Db | 670 | DETNI CNKGPVDGLTTLNGLTVAFRGHYFWMLSFPSPSPSAPRITTEVWGTPSIDIVFT | 729 |
| Qy | 1108 | RCNCEGKTTFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCIIVAALSTAKYKWPESVY | 1167 |
| Db | 730 | RCNCEGKTTFFPKDSQYWRFTNDIKDAGYKPIFKGFGGLTQCIIVAALSTAKYKWPESVY | 789 |
| Qy | 1168 | FFKRGSGIQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRRFERAIGPSQTHIRIQY | 1227 |
| Db | 790 | FFKRGSGIQYIYKQEPVQKCPGRRPALNYPVYGEMTQVRRRRFERAIGPSQTHIRIQY | 849 |
| Qy | 1228 | SPARLAYODKGLVHNEVKVSTLWKLGNVWTSALSLPNIRKPDGDYVAFSKOXYNIDV | 1287 |
| Db | 850 | SPARLAYODKGLVHNEVKVSTLWKLGNVWTSALSLPNIRKPDGDYVAFSKOXYNIDV | 909 |
| Qy | 1288 | PSRTARAITTRSGQTLKVVWVNC | 1311 |

QY 623 TLKEPAPPTPKPAPKELAPTTTKBPTSTSDKAPATPPKGTAPTTTKEPAPTTPKEPAP 682
DB 539 --KEPEPTTPKPP-----EPTTPKEPVPTTPKEPEP 567
QY 683 TTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKGPTSTSDKAPATPPKGTAPTTTKEPAPTTPKEPAP 742
DB 568 TTP-----KEPEPTTPKEPEP-----TTPKEPEPTTPKEPEPTTPKEPE 606
QY 743 PTTTPKAPPTTPETPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKE 802
DB 607 PTPPKPEPTT-----TSPKT----- 617
QY 803 PGVPTTKTAAKPEMTTAKOKTTERDLRTTPEPTTAAKPKMTKETATITTEKTESKITA 862
DB 618 -----TSPKT----- 622
QY 863 TTTQVTSITTQDTPPKITTLTKTTLAPKVTTTITTKITTTTEIMNKPDEETAKPKDRATNSK 922
DB 623 -----TTLKATTLAPKVTAPE-----EIQNKPEETTPASEDSDDSK 659
QY 923 AT-----TPKP-QKTPKAPKPTSTTKPKTMRPRVRKPTTTPTRKMTSTNPELNPT 973
DB 660 TILKPKQPKAPKPTKPKKPTKAPKPTSTTKPKT-PKTRKPKTTPAPLKTTSATPELNPT 718
QY 974 RTAEMLOTTTRNQTNSKLVENPKSEDAAGBETPHMLLRPHVFMPEVTPDMYLP 1033
DB 719 --LEVMPTTTIPKQTPNPETAENVDPDHEDADGEGEKP-LIPGPPVLFPTAIPGTDLLA 775
QY 1034 RVNNOGIIINPMSDETINICNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRIT 1093
DB 776 GRNLRGINPMSDETINICNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRIT 835
QY 1094 EWMGIPSPIDTVTRCNCEGKTFKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAA 1153
DB 836 EWMGIPSPIDTVTRCNCEGKTFKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAA 895
QY 1154 LSTAKYKNWPESVYFFKRGSGIQOYIYKQEPVQKCPGRRPALNYPVYGMTOVRRRFE 1213
DB 896 LSIKAKYKDRPESVYFFKRGSGIQOYIYKQEPVQKCPGRRPALNYPVYGMTOVRRRFE 955
QY 1214 AIGPSQTHTRIOVS-PAFLAYQDKGLHNEVKVSLWRGLPNVVTSAISLPIRKPDGY 1272
DB 956 AVGPFQTHTRIHVSVMRVSQDKGLHNEVKVSLWRGLPNVVTSAISLPIRKPDGY 1015
QY 1273 DYAFSKDQYNNIDVPSTARAITTRSGQTLISKWYNCP 1311
DB 1016 DYAFSKDQYNNIDVPSTARAITTRSGQTLISKWYNCP 1054

RESULT 6

O77765 PRELIMINARY; PRT; 401 AA.
ID O77765
AC O77765;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Superficial zone protein (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99120896; PubMed=9920774;
RA Flannery C.R., Hughes C.E., Schumacher B.L., Tudor D., Aydelotte M.B.,
RA Kuetner K.E., Caterson B.;
RA "Articular cartilage superficial zone protein (SZP) is homologous to
RT megakaryocyte stimulating factor precursor and is a multifunctional
RT proteoglycan with potential growth-promoting, cytoprotective, and
RT lubricating properties in cartilage metabolism.";

RL Biochem. Biophys. Res. Commun. 254:535-541 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular cartilage;
RX MEDLINE=99171663; PubMed=10073655;
RA Schumacher B.L., Hughes C.E., Kuetner K.E., Caterson B.,
RA Aydelotte M.B.;
RA "Immunodetection and partial cDNA sequence of the proteoglycan,
RT superficial zone protein, synthesized by cells lining synovial
RT joints.";
RL J. Orthop. Res. 17:110-120 (1999).
DR EMBL; AF056218; AAD13404.1; -;
DR InterPro; IPR000585; Hemopexin.
DR Pfam; PF00045; Hemopexin; 2.
DR SMART; SM00120; HX; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
SQ NON_TER 1
SQ SEQUENCE 401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;
Query Match 24.3%; Score 1714.7; DB 2; Length 401;
Best Local Similarity 80.2%; Pred. No. 1.3e-28;
Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;
QY 914 PKDRATNSKATTPKPKQKPTKAPKPTSTTKPKTMRPRVRKPTTTPTRK-MTSTMPBLNPT 972
DB 5 PKGRATNSQVTTTPKPKQKPTKAPKPTSTTKPKT-PRVRKPKTTPTPPKTTTSAMPBPTPT 63
QY 973 SRTAEMLOTTTRNQTNSKLVENPKSEDAAGBETPHMLLRPHVFMPEVTPDMYLP 1032
DB 64 S-LPEAMLOTTTRNQTNSKLVENPKSEDAAGBETPHMLLRPHVFMPEVTPDMYLP 122
QY 1033 PRVNOGIIINPMSDETINICNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRI 1092
DB 123 VRGSGQFGINPMSDETINICNGKPDVGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRI 182
QY 1093 TEWMGIPSPIDTVTRCNCEGKTFKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAA 1152
DB 183 TEWMGIPSPIDTVTRCNCEGKTFKDSQYWRFTNDIKDAGYKPKIPKFGGLTGQIVAA 242
QY 1153 ALSTAKYKNWPESVYFFKRGSGIQOYIYKQEPVQKCPGRRPALNYPVYGMTOVRRRFE 1212
DB 243 ALSIAQYKSPESVYFFKRGSGIQOYIYKQEPVQKCPGRRPALNYPVYGMTOVRRRFE 302
QY 1213 RAIGPSQTHTRIOVS-PAFLAYQDKGLHNEVKVSLWRGLPNVVTSAISLPIRKPDGY 1272
DB 303 RAIGPSQVHTIRIHVTPVRVYQDKGLHNEVKVSLWRGLPNVVTSAISLPIRKPDGY 362
QY 1273 DYAFSKDQYNNIDVPSTARAITTRSGQTLISKWYNCP 1311
DB 363 DYAFSKDQYNNIDVPSTARAITTRSGQTLISKWYNCP 401

RESULT 7

MUC2 HUMAN
ID MUC2 HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RA "Molecular cloning of the human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";

RL J. Biol. Chem. 269:2440-2446(1994).
 RN [2]
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=93016075; PubMed=1400449;
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
 RA Kim Y.S.;
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
 RT both upstream and downstream of its central repetitive region.";
 RL J. Biol. Chem. 267:21375-21383(1992).
 RN [3]
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
 RX MEDLINE=9358717; PubMed=1885763;
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
 RA Petersen G.M., Kim Y.S.;
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
 RT and polymorphism.";
 RL J. Clin. Invest. 88:1005-1013(1991).
 CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
 CC other mucus membrane-containing organs. Thought to provide a
 CC protective, lubricating barrier against particles and infectious
 CC agents at mucosal surfaces.
 CC -!- SUBUNIT: Multimeric.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,
 CC bronchus, cervix and gall bladder.
 CC -!- PTM: All cysteine residues are involved in intrachain or
 CC interchain disulfide bonds (By similarity).
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
 CC varies among different alleles.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 WFPC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L21998; AAB95295.1; -;
 DR EMBL; M74027; AAB59875.1; -;
 DR EMBL; M94131; AAB59163.1; -;
 DR EMBL; M94132; AAB59164.1; -;
 DR PIR; A49963; A43932.
 DR Genew; HGNC:7512; MUC2.
 DR MIM; 158370; -;
 DR InterPro; IPR002919; Cysrich_TIL.
 DR InterPro; IPR006208; Cys knot.
 DR InterPro; IPR006207; Cys knot_C.
 DR InterPro; IPR006209; EGF like_
 DR InterPro; IPR001007; WFC.
 DR InterPro; IPR001846; WFC_D.
 DR Pfam; PF00007; Cys_knot; 1.
 DR Pfam; PF01826; TIL; 1.
 DR Pfam; PF00093; WVC; 1.
 DR Pfam; PF00094; VWD; 4.
 DR SMART; SM00041; CT; 1.
 DR SMART; SM00214; WVC; 2.
 DR SMART; SM00216; WVD; 4.
 DR PROSITE; PS01185; CTCK 1; 1.
 DR PROSITE; PS01225; CTCK 2; 1.
 DR PROSITE; PS00022; EGF 1; UNKNOWN_1.
 DR PROSITE; PS01208; WFC_1; 2.
 DR PROSITE; PS0184; WFC_2; 2.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20 Potential.
 FT CHAIN 21 5179 Mucin 2.
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
 FT REPEAT 1401 1416 1.
 FT REPEAT 1417 1432 2.

FT REPEAT 1433 1448
 FT REPEAT 1449 1464
 FT REPEAT 1465 1471
 FT REPEAT 1472 1478
 FT REPEAT 1479 1494
 FT REPEAT 1495 1517
 FT REPEAT 1518 1533
 FT REPEAT 1534 1556
 FT REPEAT 1557 1572
 FT REPEAT 1573 1596
 FT REPEAT 1597 1612
 FT REPEAT 1613 1635
 FT REPEAT 1636 1651
 FT REPEAT 1652 1675
 FT REPEAT 1676 1683
 FT REPEAT 1684 1699
 FT REPEAT 1700 1715
 FT REPEAT 1716 1731
 FT REPEAT 1732 1747
 FT DOMAIN 4815 4886
 FT DOMAIN 4924 4991
 FT DOMAIN 5075 5160
 FT DISULFID 5075 5122
 FT DISULFID 5089 5135
 FT DISULFID 5098 5152
 FT DISULFID 5102 5154
 FT DISULFID ? 5159
 FT CARBOHYD 163 163
 FT CARBOHYD 423 423
 FT CARBOHYD 670 670
 FT CARBOHYD 770 770
 FT CARBOHYD 894 894
 FT CARBOHYD 1139 1139
 FT CARBOHYD 1154 1154
 FT CARBOHYD 1215 1215
 FT CARBOHYD 1230 1230
 FT CARBOHYD 1246 1246
 FT CARBOHYD 1787 1787
 FT CARBOHYD 1820 1820
 FT CARBOHYD 4339 4339
 FT CARBOHYD 4351 4351
 FT CARBOHYD 4362 4362
 FT CARBOHYD 4373 4373
 FT CARBOHYD 4422 4422
 FT CARBOHYD 4438 4438
 FT CARBOHYD 4502 4502
 FT CARBOHYD 4616 4616
 FT CARBOHYD 4627 4627
 FT CARBOHYD 4752 4752
 FT CARBOHYD 4787 4787
 FT CARBOHYD 4881 4881
 FT CARBOHYD 4888 4888
 FT CARBOHYD 4955 4955
 FT CARBOHYD 4970 4970
 FT CARBOHYD 5019 5019
 FT CARBOHYD 5038 5038
 FT CARBOHYD 5069 5069
 FT CONFLICT 1351 1351
 FT CONFLICT 1412 1412
 FT CONFLICT 1449 1449
 FT CONFLICT 1504 1504
 FT CONFLICT 4192 4192
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
 Query Match 19.8%; Score 1398.2; DB 1; Length 5179;
 Best Local Similarity 20.0%; Pred. No. 8.6e-20;
 Matches 458; Conservative 82; Mismatches 507; Indels 1248; Gaps 76;
 QY 6 LPYVLLLLSVFVIQVSSQDLSSCAGRCG-----EG-----YSR 40
 Db 502 LQVQLAPVQLFVTLDDQASQ--GQVQLCGNFGNLEGGDFKTSGLVNTAGAGFNTWKA 559
 QY 41 DATCN-----CDY--- 48

[illegible]

| | | | |
|----------|-------------|---|---------------------------|
| QY | 448 | TTKSAP--TTKPEPGTTTKEPAPTPEK-----PAPTTPKK-----PAPTTPK | 489 |
| Db | 473 | TSTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTSGPTTTTTPR | 532 |
| QY | 490 | E-----PAPTTPKEPAPTTPKKPAPTAPKE-----PAPTTPKETAPTTPKLTTP | 533 |
| Db | 533 | STTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTSGPTTTTSGPTT | 592 |
| QY | 534 | TTPEKLAPTTPKAPPTTPEELAPTTPEEPTTPPEE-----PAPTTPKAAAPNTPK | 585 |
| Db | 593 | TTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTTPRSTTTTSG | 652 |
| QY | 586 | EPAPTTPKE-----PAPTTPKE-----PAPTTPKET-----APTTPKGTAP | 621 |
| Db | 653 | GPPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTTPRSTTT | 712 |
| QY | 622 | TTLKEPAPTTPKKP-----APKELAPTTPKKEPTSTT-----SDKPAPTTPK | 662 |
| Db | 713 | TSTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTSGPTTTTSGPTTTTTPR | 772 |
| QY | 663 | GTAPTTPKSPAPTPEK-----PAPTTPKGTAPTTLKEPAPTTPK----- | 703 |
| Db | 773 | STTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTSGPTTTTSGPTTT | 832 |
| QY | 704 | PAPKELAPTTPKGTSTT-----SDKPAPTTPKETAPTTPKEPAPTTPK----- | 748 |
| Db | 833 | TTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTPRSTTTTSGPTTTTSGPTTTTTPRSTTTTSG | 892 |
| QY | 749 | -PAPTTPPET-----PPTTSEVSTPTTKEPTTIHKSPDESPPELSAEPKALENS | 799 |
| Db | 893 | CPTTTTTPRSTTTTCTSCPTTTTPRSTTTTCTSGPTT--TTPRSTTKTSTCAPT-----TTT | 946 |
| QY | 800 | PKEPGVPTTKPAATKPEMTTAKOTTERDLRTTPTETTTAAPKMTKETATTTTEKTESK | 859 |
| Db | 947 | PRSTTTTTSRPTTTTPRSTTTT--TTSRPTTTTPRSTT--TPSTSRPTTTTPRSTT-- | 999 |
| QY | 860 | ITATTQVSTTTQDTPPKITLTKTTLAPKVVTTKKTITTTTEIMNKPEETAKPKDRAT | 919 |
| Db | 1000 | -TTTSTSRPTTTTPRSTT-----KTSTCAPTTTTPRSTTT-----TTSRPTTTP | 1043 |
| QY | 920 | NSKATTPKPKPTKAPK-----KPTSTKPKTTPRVRKPKTTPTRKMT-----STMP | 967 |
| Db | 1044 | RSTTTTTPRPTTTTPRSTTTTCTSRPTTTTPRSTTTTTPRSTTTTTPRSTTTTPCPTTTP | 1103 |
| QY | 968 | ELNPTSRABAMLQTTTR-----PNQT-----PNSKLVEVN | 998 |
| Db | 1104 | SASPTR-----TTTTRRPCPCHPQPPYQIIPKSWVNPNTYENPWPQPN----- | 1149 |
| QY | 999 | PKSEDAGAGETPHMLLRPHVFMPEVTPMDYLPVNPQGIINPMLSDETNLCNGKPV | 1058 |
| Db | 1150 | -----PVPWPQMPQYFNQ----- | 1164 |
| QY | 1059 | DGLTTLRNGTLVAFRGHYFWMLSPEPSPARRITEVWGIPSPIDTFTVTRCNCBGT | 1115 |
| Db | 1165 | -----WPQLGYPQOLPPP-----LPSP----- | 1182 |
| QY | 1116 | FFPKDSQVWRFTNDIKDAGYKPK-----IFKFGGLTGQIVAAALSTAKYXN | 1161 |
| Db | 1183 | -----QMPWS-----WPKFPVQPGDCENICNLKGK----- | 1211 |
| QY | 1162 | WPESVYFFKRGSGTQOYTYKQEPVQKC | 1188 |
| Db | 1212 | -----YQEDLIRRC | 1220 |
| RESULT 9 | | | |
| Q6DNC4 | ID | Q6DNC4 | PRELIMINARY; PRT; 251 AA. |
| AC | Q6DNC4 | Q6DNC4 | Q6DNC4 |
| DT | 01-OCT-2004 | (TREMBlrel. 28, Created) | |
| DT | 01-OCT-2004 | (TREMBlrel. 28, Last sequence update) | |
| DT | 01-OCT-2004 | (TREMBlrel. 28, Last annotation update) | |

Db 1031 TLKTEBPTTRKSTTKTTREPTTRKRVTTERTTREPSTKTTTHHTAEPATKTTHEP 1090
 QY 659 TTPKGTGTA-----PTTPKEPAPTTPKEPA-----PTTPKGTAPTTLKPEPA--- 697
 Db 1091 TTOKSTTLRITBEPITRKSSTAKTTRPTTKRETTERTTKEPTTRKTTTHKTTBEPPTTK 1150
 QY 698 -----PTTPK-----KPAPKELAPTTPKGTSTTSKPAPTTPKETAPTTPKEPA----- 742
 Db 1151 TTHTBPTTKSTTLKPTEE---PTTRK---TSTTKTTRPTTKKTTERTTKEPTTRKTTTH 1205
 QY 743 -----PTTPK-----KP-----APTTPKPPPTTSVSPPTTK 771
 Db 1206 KTBEBPTTNKTKTKTTEPTTKKSTLAPTBEPTTRKSTTKTTRPTTKRVTTERTTR 1265
 QY 772 EP-----TTHKSPDE-----STPELSABPTPKALENSKPEGVPTTPKPAATKPMPTT 821
 Db 1266 EPTTRKTTTHKTEBPTTKTKTKTTEPTTKKSTLAPTBEPTTRKSTTKTTRPTTKRVTTERTTR 1324
 QY 822 AKDKTERDLR-----TTPETTTAAP----- 842
 Db 1325 TKRVTTERTTRBPTTRKTTTHKTEBPTTKTKTKTTEPTTKKSTTLKPTBEPTRKT 1384
 QY 843 ----KMTKETAT---TTEKTESKITATTQVTSITTTQDTPPKITTLKT-----TT 887
 Db 1385 STTKTTRPTTRKRVTTERTTREPTRKT---THKTTBPTTKTKTKTTEPTTKKST 1442
 QY 888 LAP-----KVTTK-----KITTTTEIMNKP-----BETAKPKORATNSKATT 925
 Db 1443 LKTEBPTTRKSTTKTTRPTTKRVTTERTTREPTRKTTHKTTBPTTKTKTKT 1502
 QY 926 PKQO-----KP-----TKAPKKPT----- 939
 Db 1503 HEPTTKKSTTLKPTBPTTRKSTTKATREPTTKRVTTERTTREPTRKTTHKTTBPTTK 1562
 QY 940 -----STKKPKTMRVRPKPTTPTRKMTSTPELNPTSRIAEAMLQIT--- 983
 Db 1563 TKTKTKTKTTEPTTKKSTTL-----KTEBPTTRKSTTKTTRPTTRKSTSVKITADQ 1618
 QY 984 -----TRPNQTPNSKLAVNPKSDEAGAGETP-----HMLRPHV 1020
 Db 1619 TTKRTTAEMSTTNOEPTSVTTNSSQNTTSTTSTTTEBQVHHHHHHIHYH---KPAD 1675
 QY 1021 FMEVTPDMYDLPRVFNQGLIINPLMSDEINI CNKPKVDGLTLNGLTVAFRGHYFWML 1080
 Db 1676 LGPSILP-----LPDLP----- 1687
 QY 1081 SPSPSPSPARRITEVWGIPSPIDTVTRCNCKGKTFPFKDSQYWRFTNDKAGYKPIF 1140
 Db 1688 -PLPLPLP-----WP-PLPLPEI-----PLPL- 1707
 QY 1141 KGFGLTGTGVAALSTAKYNWPSVYVFKRGSGIQIYVQBPVQKQGRPALNVPVY 1200
 Db 1708 -----PPLPTA-----LPPLPLPLPLP----- 1725
 QY 1201 GEMTOVRRRFRFAIGFSQTHTRIQYSPARLAYQDKVLHNVKYSILWGLNPNVTS 1260
 Db 1726 -----LPEVNLT 1733
 QY 1261 ISLPNIRKPD 1270
 Db 1734 ISLPEISLPN 1743
 RESULT 11
 ID AAS64673 PRELIMINARY; PRT; 1761 AA.
 AC AAS64673;
 DT 01-APR-2004 (TrEMBLrel. 27, Created)
 DT 01-APR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 01-APR-2004 (TrEMBLrel. 27, Last annotation update)
 DE CG33300-PA (Fragment).
 GN CG33300.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer V.C., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,
 RA Khril J.F., Aqbayani A., An H.J., Andrews-Pfannkuch C., Baisley E.M.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Brothier P.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Hernandez J.R., Houck J.,
 RA Harris N.L., Harvey D., Heiman T.J., Howland T.J., Wei M.H., Ibegwam C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Talali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Patterson K., Saunders R.D., Scheeler F., Shen H.,
 RA Reiner B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RL "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
 RL "Finishing a whole-genome shotgun: release 3 of the Drosophila
 melanogaster euchromatic genome sequence."
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RL "Annotation of the Drosophila melanogaster euchromatic genome: a
 systematic review."
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [4]
 RP SEQUENCE FROM N.A.

RA MEDLINE-22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Swirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.B., Rubin G.M.,
RA Ashburner M., Celisner S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [5]
RP SEQUENCE FROM N.A.
RF FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RF FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE003626; AAS64673.1; -
FT NON TER
SQ SEQUENCE 1761 AA; 197521 MW; 62A5E1656241B3F4 CRC64;

Query Match 18.6%; Score 1311.6; DB 2; Length 1761;
Best Local Similarity 23.1%; Pred. No. 7e-19;
Matches 459; Conservative 108; Mismatches 379; Indels 1044; Gaps 88;

QY 74 CF-----ESFERGRCDDCAOCKYDKC-----CPDYESFOAEVKN 110
DB ||||| : : : : : ||
D5 CFIIWALXHSANAALLEGSEIRFGDESITPESAKEYPDYCWINPFLPGCP----- 52
QY 111 KKNRTKKPKTPKPPVWDAGSLGDGPKVTTPDTSTTQHNVKVSFKIIT----- 161
DB ||||| : : : : : ||
D53 -----FPGGGNNSTTKISLATIKTSVTSTBEETTLKTIIKST 92
QY 162 AKPNRPSLPNSDTSKETSLTVNKETVETKETTNTKOTSDGKE---KITSAKET 217
DB ||||| : : : : : ||
D93 AKPTQKTNEPTTE---KITTPKATTI--KSTATTARATEAPKTEQTLRTTIKST 145
QY 218 Q----SIEKTSAKDLAPTSQLAKPTPKAETTTKGPAALTTPKEPT-----PTT 261
DB ||||| : : : : : ||
D146 SELTILUKTITIKSTAETKSKTHNPTTKSTT-----LRTEPTTRKSSTAKTTREPT 200
QY 262 KPASTTTPKEPT-----PTTIKSA-----PTTPKEPAPTTT 293
DB ||||| : : : : : ||
D201 KRETTERTTQEPSISKITTHTTAEPATKKITHEPTTOKSTTLRIETEPTRKSSAKTT 260
QY 294 KSAPTTPEKPAPTTTKEPA-----P312
DB : ||||| : : : : : ||
D261 RE-PTTKRETTERTTKEPSTKITTHTTAEPATKKITHEPTTOKSTTLRIETEPTRK 319
QY 313 -----PTTPKEPAPTTTKEPA-----P329
DB ||||| : : : : : ||
D320 STARTTREPTTKRETTERTTQEPSSTKITHETTAEPATKKITHEPTTOKSTTLRIETEPT 379
QY 330 TTTKSA-----PTTPKEPAPTTPKKPA-----PTTPK-----356
DB ||||| : : : : : ||
D380 TURKSSTAKTTREPTTKRETTERTTKEPSTKITTHTAAEPATKKITHEPTTOKSTTLRI 439
QY 357 ---EPA-----PTTPKEPTPTTKEPA-----PTTKEPA-----PTTPK 387
DB ||||| : : : : : ||
D440 ITSEPTTRKSSTARITREPTTKRETTERTTQEPSSTKITHETTAEPATKKITHEPTTOK 499
QY 388 -----EP-----APTAPKKPA--398
DB ||||| : : : : : ||
D500 STTLRIETEPTTRKSSTAKTTREPTTKRETTERTTQEPSSTKITTHTTAETAATKKITHT 559
QY 399 PTPPK-----EPA-----PTTPKEPAPTTTKEPS-----422
DB ||||| : : : : : ||
D560 PTTCSTTLRIETEPTTRKSSTARITREPTTKRETTERTTQEPSSTKITTHTTAEPATKK 619
QY 423 ----PTTKEPA-----PTTKSA-----PT-----TTK 442
DB ||||| : : : : : ||
D620 KTTHEPTTOKSTTLRIETEPTTRKSSTARITREPTTKRETTERTTQEPSSTKITTHTTA 679
QY 443 EPA-----PTTKSA-----PTTKEPSPTTKEPA-----468
DB ||||| : : : : : ||


```

Db 1652 RTTPTPTPTDMMSSASTPEPTTPTPTPTPTDSTMMSSMSSESTPEPTTPE 1711
Qy 759 PTT-----SEVSTTTTKEPTTIHKSP-----DE 782
Db 1712 PSITPGTTRTPTPTSTESTDMMSSASTPEPTTPTPTPTPTSTESTDMMSSA 1771
Qy 783 STPELSAEP-----TPKALENSKEP-----GYPTT 808
Db 1772 STPESTPTGTRTPTPTPTSTESTDMMSSASTPEPTTPTPTPTPTSTESTDMM 1831
Qy 809 KTPAATKPEMTT-----TAKDKTTTERDLRITP-----835
Db 1832 MSSASTPEPTTPTPTPTPTPTPTSTESTDMMSSASTPEPTTPTPTPTPTSTES 1891
Qy 836 -----ETTTAAKMKIKETATT-----TEKTTESKIT 861
Db 1892 TDTMMSSASTPEPTTPTPTPTPTPTPTSTESTDMMSSASTPEPTTPTPTPTPT 1951
Qy 862 ATTQVISTTQD-----TTPFKIT-----TLKTTTLAP 890
Db 1952 PTETMMSSASTPEPTTPTPTPTPTPTPTPTSTESTDMMSSASTPEPTTPTPT 2011
Qy 891 KVTTTKTITITELMKPBTAKP-----KDRATNSKATTPKQ-----929
Db 2012 --TSTERTDMMSSASTPEPTTPTPTPTPTPTPTPTSTESTDMMSSASTPEPTT 2069
Qy 930 ---KPTK-----APKPTSTK-----KKTMPRVR 951
Db 2070 TPTPTPTSTDSMMSSASTPEPTTPTPTPTPTPTPTPTSTESTDMMSSASTPE 2129
Qy 952 K-----PKTTPPKMTSTMPELNPT-SRIAEAMLQTTTP 986
Db 2130 RTTPTPTSTDSMMSSASTPEPTTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2187
Qy 987 NOTPNSKLVNPKSEDAGAGETHMLLRPHVFMPEVTPDMDYLPRVENQGIINPML 1046
Db 2188 GTTPTPTPTPTSTESTDMMSSASTPE-----PSTTPG-----2220
Qy 1047 SDETNICNGKPDGLTLRLANGTLVAFRHYFWMLSPFPSPA-----RRTTEVWGI 1101
Db 2221 TTRTPTPTPTPD--STMSS-----MSSASTPEPTTPTPTPTPTPTPTPT--RPT 2264
Qy 1102 IDTVFRNCEGKTFEFDKSOYWRFTNDIKDAGYKPIKFGGLT-----1147
Db 2265 TDTT-----MSSMSASTPEP--STPTGTTTPTPTPTPTPTPTPTDSTM 2302
Qy 1148 GOIVAAALSTAKYNWPESVYFFKRGGSIOQ-----YIYKQB- 1183
Db 2303 SSSMSASTPESSITPE-----LQDQHLPIQCRLVCROFLLRSHRHLVQPEL 2353
Qy 1184 -----PVQKCPGRPALNYPVYGMQVRRRFRERALGPSQTHIRIQYSPALAYQ 1236
Db 2354 LQDQHLPIQC--RQPL-----RSHLRHLVQP-----ELQD 2386
Qy 1237 KGVLHNEVKVILMRLGNLVVT--SAISLPIRKPQDGYDYAFKSDQYINIDVRSARA 1294
Db 2387 -----QHLFSTDTMMSSASTPE-----PSTTPG- 2409
Qy 1295 ITRRSQT 1302
Db 2410 -TTRTPT 2416

```

RESULT 13

```

Q6SSE6 PRELIMINARY; PRT; 3409 AA.
AC Q6SSE6;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE plus agglutinin.
GN Name=SAG1;

```

```

OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadales; Chlamydomonas.
OX NCBI_TaxID=3055;
RN 11 SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBAJ databases.
DR EMBL; AY450930; AAS07044.1;
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01218; PSTEXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FB87633EDD6631F CRC64;

Query Match 17.28; Score 1211.6; DB 2; Length 3409;
Best Local Similarity 13.74; Pred. No. 3.2e-16;
Matches 384; Conservative 167; Mismatches 447; Indels 1814; Gaps 78;

Qy 1 MAWKTLPI-----YLLLLLSV-----16
Db 50 LAWTTIPVSDALQFPVNVSVTSSAATDAPSAIAYIGALLDLVNWVQPCINDTRSYD 109
Qy 17 -----FVIQYSS-----ODLSSACRGCGEYSRD 41
Db 170 DFGIYQPVLDGDFMMWGFVTNTVDSNLKWIDPAWTAQGAWLGGDF-----RD 217
Qy 42 ATCN-----CDYNCQHYMECCPDFKRVCTAELSCKRCFESFERGECDCDAOCKYD-- 94
Db 218 AVNMFQTHYCSW--PFVEC-----SSCENYDIA 243
Qy 95 -----94
Db 244 DPYDPDKIANGIVPAVITALDFPNASLIYIYLDGFMFAGSLDNWYMLNAYNFTGGPLP 303
Qy 95 -----KCCPD-----YESFCAE-----106
Db 304 ANLPTLLPSLOHLAHDHCRATEDVRGTASLOYWQOYPTSTGOPYE--YCSGDGVTGSDGTE 362
Qy 107 -----106
Db 363 YVISGMIPDENGDAVAAASDSSLFWANLRTVRLSNQALYGIPEGLRSASSISSWRILQ 422
Qy 107 -----VKDNKKNRK-----116
Db 423 GNTELCGLPEFAAPINSLLYLGLTGTWVWHPADHTQCGECLAPPPSPSPRPP 482
Qy 117 -----116
Db 483 RPPPLPPSPPPPLPPSPVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 542
Qy 117 -----KKTEKPPVVDVDEAGSLDNGDKVTTTDDT 145
Db 543 GGI CDSPSPMPSPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 589
Qy 146 ---STQHNVKYSTPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKTTTNNKQ 202
Db 590 PPASSVPPSPPEPPSPKPPSPAPPSPAPPSP-----620
Qy 203 TSTDGKEKTTSAKETQSIKETSADKDLAPTCKVLAKPTPKAETTKGPAITTPKEPTPT 262
Db 621 -----PPSPAPPSPAPPSP-----BAPSPQPPSPVP 647
Qy 263 KEPAST-----TPKEPTPTTKSAPTTPKEPAPTTTKSAPTTKSPAPTTKEPAPT 315
Db 648 QPSPVPPSPKPPSPAPPSPVPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 707
Qy 316 PK-----EPAPTTTKEPAPTTTKSAPTTTPKBPAPTTTPKKPAPT-----TPKE 357

```



```

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadales; Chlamydomonas.
OX NCBI_taxid=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RT "plus and minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AA507044.1; -
SQ SEQUENCE 3409 AA; 336045 MW; 7FB87633EDD6631F CRC64;

Query Match 17.2%; Score 1211.6; DB 2; Length 3409;
Best Local Similarity 13.7%; Pred. No. 3.2e-16;
Matches 384; Conservative 167; Mismatches 447; Indels 1814; Gaps 78;

QY 1 MAWKTLPI-----VLLLLLSV----- 16
Db 50 LAWVTIPVSALQFPVNVSVTSSAATDAPSAIAYIGALLDLVWVWPQNCINDTRSYD 109
QY 17 -----FVIQVSS-----QDLSSCAGRCGEGYSRD 41
Db 170 DFGIYQPVLDGFMWVGFTVNTVDSNLKWI DPATQAQGLGQDF-----RD 217
QY 42 ATCN-----CDVNCQHYMECCDFKRVCTAELSCKRCFESFERGECDCDAOCKYD-- 94
Db 218 AVWNEFTHYCSW---PFVEC-----SSCENYDIA 243
QY 95 -----KCCPD-----YESFCAE----- 106
Db 304 ANLPTLPSLOHALDHCRATPDVRGTASLQYGMQYPTGTQPYE-YCSGDVGTGSDGTE 362
QY 107 -----KCCPD-----YESFCAE----- 106
Db 363 YVISGMIPEWGDVAANAASDSSLWANLRTVRLSNQALYGIPEGLRSASSISSWRLQ 422
QY 107 -----VKDNKKNRK----- 116
Db 423 GNTELCGLPEFAPIINSLLYGLTGHVWHPADHTQGECLAPPPPPSPRPP 482
QY 117 ----- 116
Db 483 RPPPLPSPPPPLLPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 542
QY 117 -----KKPTKPEVVDVDEAGSLDNGDFKVTTPDT 145
Db 543 GGICDPSMPSPRPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPSPDS 589
QY 146 ---STQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSLVNKTETVETKETTINKQ 202
Db 590 PPASSVPPSPPPSPKPPSPAPPSPAPPSPPP----- 620
QY 203 TSDGKEKTSKAKTSIEKTSKADLAPTSKVLAKPTPKAETTTKGPAITTPKEFTPTTP 262
Db 621 -----PPSPAPPSPAPPSP-----PAPPSPQPPSPVPP 647
QY 263 KEPAST-----TPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTT 315
Db 648 QPPSPVPPSPKPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 707
QY 316 PK-----BPAPTTKPEAPTTTKSAPTTKPEAPTTTKPKKAPT-----TPKE 357
Db 708 PSPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 767
QY 358 PAPTTPKEPTPTTKPEAPTTKPEAPTTKPEAPTT-----APKKPAPTTPKPEAPTTKPE 413

```

```

Db PAPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 827
QY 414 APT-----TTKEPSPTTPKPEAPTTTKSAPTTKEP 444
Db 828 SPAPPSPPPSPQPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 887
QY 445 APTTTKSAPTTKPEPSTTTKPEAPTTTKPEAP-----TTPKKAPATTPK 489
Db 888 SPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 946
QY 490 EPAPTTKPEAPTTTKKAPATPK-----EPAPTTTKETAP 525
Db 947 PPSPPPPPSPAPPSPAPPSPPTTTPSPVPSPAPPSPAPPSPAPPSPAPPSPAPP 1006
QY 526 TTPKLTPTTTPKELAPTPPE--KPAPTTPEELAPT-----TPPEPTTTP 569
Db 1007 NPSPVPPTTTPSP--GPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1065
QY 569 EPAPTTTKAAAPN-----TPKEAPTTTKPEAPTTK-----EPAPTTTKETAPTKG 618
Db 1066 EPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPP 1125
QY 619 TAPTTL-----KEPAPTTTKKAPKELAP 642
Db 1126 PSAPLPPSPDPSPAPPSPAPPSPMPLPTSPSPSPVPPTTTPSPAPPSPAPPSP 1185
QY 643 TT-----TKBPTSTSDKAPATTTPKGTATTPKPEAPTTTK 678
Db 1186 PSPEPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1244
QY 679 EPAPTTKGTATTTKPEAPTTTKKAPKELAPT--TTKGPTSTTSKAPATTTPKETAPT 737
Db 1245 PPSAPPSPQPPS-----PAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 1297
QY 738 PKEAPTTKPEAPTTTPETPPT--TSEVSTPTTTKPEPTTIHKSPDESTPELSAETPKAL 796
Db 1298 PEPSPAPPQPPSPVPSPAPPSPPTTTPAPAPAAALPPLPSPAPPPLVPVPSPAPP 1357
QY 797 E-----NSPKEGV---PTTKTP----- 811
Db 1358 RPPQPTAMPPPSPAPPSPAPPSPAPPSPAPPSPPPPTTTPPLAPLPPDCTLLAQALLSIP 1417
QY 812 -----AATKPE-----MTTAA----- 822
Db 1418 DAANSFVVSAGLPTSVAPSTPPELLASFCTVCSQLTATAISLVGSGSRGNTNGSSG 1477
QY 823 -----KDKTTERDL 831
Db 1478 GNYISNGDDAAIQRIIPAGNSTDROGSSSGSGSGSWGPGTETAABWAVDAVDGTYQL 1537
QY 832 -----RTTP----- 835
Db 1538 SIGGVYTRTVVDRTPPSVSGNVTLSNRKIQEPSAVGEASLNALGSKQAMLLTIFSE 1597
QY 836 -----ETTTAAPKMT-----KETATTTEKTESKIT----- 861
Db 1598 PVPAPDPAASLIIVTGALVAEWAAADKMTFVYLAWTLPAELVATAAGSSSSGTSRSGN 1657
QY 862 ----- 861
Db 1658 NGTAAAAAAPPAGTTGRRRALQOAAAAAPPSPASGSSSSLSGAATAAQQRHVHFL 1717
QY 862 -----ATTQVSTSTTQDTTPPKITLK 884
Db 1718 PATAYADAARNPGENDLSLVELTDNAVSPAVGEALATTARVTAATYP-----AVAA 1770
QY 885 TTTLAPKVT----- 893
Db 1771 TTTLVAASSSFAAIRAKGSLQGSVHIQMLTMSLVASRGVGREYGEYAVBFKYAVLG 1830
QY 894 -----TTKKTITTE-----IMNKEEFTAKPKDRATNSKAT 924

```

SEQUENCE FROM N.A.
Jones A.R., Hughes C.E., Flannery C.R., Caterson B.;
Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
EMBL; AY536119; AAS83977.1; -;
InterPro; IPR000585; Hemopexin.
Pfam; PF00045; Hemopexin; 2.
SMART; SM00120; HX; 2.
PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
NON_TER 251 251
NON_TER 251 251
SEQUENCE 251 AA; 28532 MW; 2011B2C1719AA68B CRC64;
Query Match 16.7%; Score 1183; DB 2; Length 251;
Best Local Similarity 87.2%; Pred. No. 8e-18;
Matches 218; Conservative 10; Mismatches 22; Indels 0; Gaps 0;
QY 1055 KPVVDGLTLRNGTLVAFRGHYFVWMLSPFSPSPARRITEVWGIPSPIDTVEVTRCNCEGK 1114
Db 1 GRPVDGLTLRNGTLVAFRGHYFVWMLSPFSPSPARRITEVWGIPSPIDTVEVTRCNCEGK 60
QY 1115 TPEFKDSQVWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFKRGGS 1174
Db 61 TPEFKGSQVWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYFFKRGGS 120
QY 1175 IQQYIYKQEPVQKCPGRRPALNYPVYEMTQVRRRFFERAIQPSQTHIRIOYSPARLAY 1234
Db 121 VQOQYTYKQEPVQKCPGRRPALNYPVYEMTQVRRRFFERAIQPSQTHIRIOYSPARLAY 180
QY 1235 QDKGVLHNEVKVSIILWFGLPNTVTSALSLPNIRKPDGYDYAFSKDQYNNIDVPSRTARA 1294
Db 181 QDKGVLHNEVKVSIILWFGLPNTVTSALSLPNIRKPDGYDYAFSKDQYNNIDVPSRTARA 240
QY 1295 IITRSGQTLS 1304
Db 241 IITRSGQTLS 250

Search completed: October 13, 2004, 12:20:38
Job time : 182.144 secs

1831 VKGNLGPAAEAMPTNEKEVTAAEQARQVGGDLWPIGNDLLGGSSNTTASGSSSSSSSSN 1890
QY 925 TPKEQKP-----TKAPKKPTSTK 942
Db 1891 SP-PRRPPPPAAGSTGLFSNADASPPPLAVATPALPAPLSTIAAATAAP-----1942
QY 943 KPEKTPVRKPKTKTTPPRKMTST--MPELN-----PTSRIRAEAMLOT 982
Db 1943 -----PRLPSP---PPPAVGSSTGLVPRRHLMQMLQPPFAAFAVAPPPPPASSALVLP 1994
QY 983 TTRPQNTNSKIVE-----VNPKS 1001
Db 1995 SPPPPPPSLLIQOASATYVSDMDLTYLVAAAMLITAVAAAGRLIAAVALYRLVSE- 2053
QY 1002 EDAGAGCEGTHMLLRPFVMEVETPMDYLPVRPNQOIIINPMLSDETNICNGKPVQGL 1061
Db 2054 -----PPLFLA-----PRLP-----2064
QY 1062 TTRNGTLVAFRGH-----YFWMLSPFPSPSPA 1089
Db 2065 TTIAGLLIIVALTFFSCMALGSPAADHMGSRATAAYCVLTIIVVYAAFLWLA-----2116
QY 1090 RRIEVMGTP-----SPIDTVTRCN-----1110
Db 2117 --LARAAMVPQTLVEPMTTSSVASPRPSAFERADRTSPRSTVSIATAADIGGTATGG 2174
QY 1111 -----CEKTEFFKDSQV---WR-FTNDIKDAGYKPIFKGFG 1144
Db 2175 GGAASACADVGMAAPATAAGADVPSEDDRYARGPHWKQF-----DGLPPTVASGVG 2229
QY 1145 GLUTQIVAAALSTAKYKNWPESVYFFKRGSIQQYIYKQEPVQKCPGRRPALNYP-----1198
Db 2230 GSGGGVVPL-----PPLVALPGSGRHLPLPPLPAA 2260
QY 1199 -----VYGMTQVRRRRPERAIGPSQTHIRIOYSPARLAYQDKG 1238
Db 2261 GAATAGGGAATAPGPRGAAGB-----DTH-----QFCP-----2291
QY 1239 VLHNEVKVSIILW-----1251
Db 2292 --H-----WRKFSVPDDGAATAGVVDAAAGAGGGGGRSGSTGDFRANGTRT 2341
QY 1252 -----GLP-----NVVT 1258
Db 2342 PSDDGAKGRSSSHGGGNGNMSGAAASFCASAGGGGAAAAASRLGSRSGSPGCDYNAMS 2401
QY 1259 SAISL-----1263
Db 2402 SGAMRGGGAGASRFAGSAGSPGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2461
QY 1264 -----PNIRKP-----DGYYAFSKDQYNNIDV 1287
Db 2462 GSGSNEGGRQMSYGSNNMTAGAVGGVRRPAGSLRNASDG-----DA 2507
QY 1288 PS-----RTARAI-----TTRSG 1300
Db 2508 PPLPAPGTVPLPFANLLTSGADYRRLNAEAVAAAGGGGGGGTSSG 2559

RESULT 15
Q6QJF3 PRELIMINARY; PRT; 251 AA.
AC Q6QJF3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Proteoglycan 4 (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactylia; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 26.7037 Seconds
(without alignments)
4723.689 Million cell updates/sec

Title: SEQ1-H
Perfect score: 7064
Sequence: 1 MAWKLPYVILLLSVFIQ.....ARAITRSGQTLSKVWVNC P 1311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | Match | Length | DB | ID | Description |
|------------|--------|---------|-------|--------|--------|----|---|
| 1 | 1382.7 | 19.6 | 3020 | 2 | A43932 | | mucin 2 precursor, intestinal - human (fragments) |
| 2 | 1154.2 | 16.3 | 1664 | 2 | T18262 | | S-layer protein - hu |
| 3 | 1038 | 14.7 | 7962 | 2 | I38346 | | elastin titin - hu |
| 4 | 1026.6 | 14.5 | 3570 | 2 | T45025 | | mucin MUC5B, trach |
| 5 | 1021.4 | 14.5 | 2187 | 2 | T30826 | | nascent polypeptid |
| 6 | 1011.3 | 14.3 | 1489 | 2 | T31108 | | cyst germination s |
| 7 | 1010.5 | 14.3 | 1274 | 2 | T16251 | | hypothetical prote |
| 8 | 956.2 | 13.5 | 1367 | 1 | S48478 | | glucan 1,4-alpha-g |
| 9 | 951.7 | 13.5 | 3507 | 2 | T34513 | | hypothetical prote |
| 10 | 945 | 13.4 | 6642 | 2 | T29757 | | protein UNC-89 - C |
| 11 | 900.4 | 12.7 | 1188 | 2 | S49915 | | extensin-like prot |
| 12 | 895.9 | 12.7 | 4135 | 2 | T42629 | | tenascin-X - bovin |
| 13 | 885.5 | 12.5 | 5762 | 2 | A18199 | | proline-rich pepti |
| 14 | 870.7 | 12.3 | 1229 | 2 | T25697 | | hypothetical prote |
| 15 | 867.5 | 12.3 | 1344 | 1 | A35175 | | mucin 1 precursor, |
| 16 | 841.9 | 11.9 | 5262 | 2 | T03454 | | ALR protein - huma |
| 17 | 837.8 | 11.9 | 2897 | 2 | B48666 | | cell proliferation |
| 18 | 837.8 | 11.9 | 3256 | 2 | A48666 | | cell proliferation |
| 19 | 824.6 | 11.7 | 3942 | 2 | T42730 | | Bassoon protein - |
| 20 | 823.4 | 11.7 | 1151 | 2 | T18535 | | high molecular mas |
| 21 | 821.3 | 11.6 | 2232 | 2 | T34434 | | hypothetical prote |
| 22 | 814.8 | 11.5 | 1832 | 2 | T31113 | | mucin-like glycopr |
| 23 | 809.4 | 11.5 | 3938 | 2 | T42761 | | Bassoon protein - |
| 24 | 808.9 | 11.5 | 2142 | 2 | B35038 | | MHC class III hist |
| 25 | 805 | 11.4 | 4957 | 2 | T03455 | | ALR protein - huma |
| 26 | 800.6 | 11.3 | 4548 | 1 | S00657 | | apoptosis(a) (EC |
| 27 | 798.7 | 11.3 | 1872 | 2 | S36152 | | MHC class III hist |
| 28 | 793.3 | 11.2 | 1870 | 2 | S37671 | | MHC class III hist |
| 29 | 787.2 | 11.1 | 4006 | 2 | T09070 | | probable tenascin |

ALIGNMENTS

RESULT 1

A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004
C;Accession: A49963; A45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002; PMID:8300571
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU>
A;Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:I21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A;Reference number: A45106; MUID:93016075; PMID:1400449
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A;Reference number: A43932; MUID:91358717; PMID:1885763
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence incontinent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A;Reference number: A33332; MUID:89197956; PMID:2703501
A;Accession: B33332
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

| | | | | | | |
|----|-------|------|------|---|--------|--------------------|
| 30 | 783.2 | 11.1 | 2225 | 2 | T26063 | hypothetical prote |
| 31 | 778.9 | 11.0 | 3381 | 2 | T42389 | versican precursor |
| 32 | 772 | 10.9 | 1630 | 2 | A53577 | ascites sialoglyco |
| 33 | 765.7 | 10.8 | 1611 | 2 | T38236 | hypothetical prote |
| 34 | 758.8 | 10.7 | 3566 | 1 | A40701 | tenascin-X precurs |
| 35 | 756.1 | 10.7 | 5105 | 2 | T32650 | hypothetical prote |
| 36 | 752.6 | 10.7 | 1777 | 2 | T34369 | hypothetical prote |
| 37 | 751.2 | 10.6 | 761 | 2 | C84672 | hypothetical prote |
| 38 | 749.2 | 10.6 | 839 | 2 | F75518 | hypothetical prote |
| 39 | 748.2 | 10.6 | 2774 | 2 | A43359 | microtubule-associ |
| 40 | 747.4 | 10.6 | 3149 | 1 | Q0888 | BLF1 protein - hu |
| 41 | 746.3 | 10.6 | 924 | 2 | S27923 | gene lF3 protein - |
| 42 | 741.7 | 10.5 | 990 | 2 | I51618 | nucleolar phosphop |
| 43 | 735.3 | 10.4 | 5170 | 2 | T15348 | hypothetical prote |
| 44 | 735.1 | 10.4 | 971 | 2 | T19431 | hypothetical prote |
| 45 | 725.8 | 10.3 | 4667 | 2 | T20774 | hypothetical prote |

J. Clin. Invest. 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481; PMID:1985113
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
R:Yu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminal domain.
A:Reference number: PQ0328; MUID:92198477; PMID:1550588
A:Accession: PQ0328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M86523
A:Experimental source: small intestine
A:Accession: PQ0329
A:Molecule type: protein
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C:Genetics:
A:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMIM:158370
A:Map position: lip15.5-lip15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
C:Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

Query Match 19.6%; Score 1382.7; DB 2; Length 3020;
Best Local Similarity 18.5%; Pred. No. 2.9e-21;
Matches 482; Conservative 126; Mismatches 567; Indels 1433; Gaps 89;

QY 6 LPYLLLLSVFVIQVSSQDLSSCAGRCG-----EG-----YSR 40
DB 502 LQVQLAPVQLFVTLQASQ--QVQGLCGNFGLEGDDPKTASGLVEATGAGFANTWKA 559
QY 41 DAFN-----CDY---48
DB 560 QSTCHDKLDWLDPCSLINIESANYAEHWCSSLKKTETPPGRCHSAVDPAEYKRCYDTC 619
QY 49 NCQHYMEC---CPDFKRVCTA-----ELSK 71
DB 620 NQNNEDCLCAALSSVAACTAKGVLWGHVHCKNDKGVSCFNSQVFLNLTCCQTCR 679
QY 72 -----GRCPESF-----ERGR-----ECDC-----DAQCKYDK 95
DB 680 SLSEADSHCLEGAPVDGCGCPDHTFLDEKRCVPLAKCSYHRGLYLEAGDVVVRQBER 739
QY 96 C-----96
DB 740 CVCRDGRHLHCRQIRLIQGSCTAPKIHMDCSNLTALATSKPRALSCOTLAAGYHTECVSG 799
QY 97 --CPD-----YESFC-----AEVK-----108
DB 800 CVCEDGLMDGGCGVVEKECPVHNDLKYSSGAKIKVDNCTCTCKGRWVCQAVCHGT 859
QY 109 -----DNK-----111
DB 860 CSYGSGHYITFDGKYVDFDGHCSYAVQVQYCGQNSLSGFSIITENVPCGTTGVTCSKA 919
QY 112 -----KNRTKKPTKPPV-----VDEAGSL-----133
DB 920 IKIFMGRTELKEDKRVVIQRDEGHVAYTREVQVLYVVESTGIIVWKRTTFEIK 979
QY 134 -----DNGDFKV-----TPDFTSTTQ-----149
DB 980 LAPSYKGTVCGLGNFDRHNSNDFTRDHMVSSSELDGNSWKEAPTCPDVSTNPEPCL 1039
QY 150 -----HNKVTSP-----157
DB 1040 NPHRRSWAEKQCSILKSSVFSICHSKVDPKPYEACVHDSCSDTGGDCFCFSAVASYA 1099
QY 158 -----KITAKPIN-----166

DB 1100 QECTKEGACVFWRTDLCIFCDYVNPHECEWHYPCGNSRFETCRTINGIHSNISVSY 1159
QY 167 -----PRPSLP-----172
DB 1160 LEGCVPRCPKDRPIVEDLKKCVTADKCGCYVEDTHYPGASVPTTEETKSCVCVNSSQV 1219
QY 173 -----PNSDTSKE-----TSLTVNKET-----TVRTK 194
DB 1220 VCRPEGKILNQTDGAFYWEICGPNGTVEKHFNCSITTPSPSLTITFTITLTPPTS 1279
QY 195 ETTTINKOTSDGKEKITSK-----ETOSIEKT 223
DB 1280 FTTTITTTTPTSSVLSLSTPKLCLLWSWINEHPSSGSDGDRPFPGVCGAPEDIECR 1339
QY 224 SAKD-----LAPT 231
DB 1340 SVKDPHLSLEQHGQKQVQCDVSVGFICKNEQDQNGNPFGLCYDYKIRVNCWMDKCIITP 1399
QY 232 SKVLAKPTPKABETTTKGPALATTPKEPTPTTPKEPASTTPKEPTTIKSAP--TTPKEPAP 290
DB 1400 SPPTTTPSPPTTTTTLPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSPPTT 1459
QY 291 TTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 341
DB 1460 TTTTTPPTTTPSP--PTTTPSPPTTTPSPPT--TTTTTPPTTTPSPPTTTPPTTTP 1517
QY 342 PAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 401
DB 1518 PTTTTPSPPTTTTTPPTTTPSPPTTTTTPSTTTTTPPTTTPSPPTTTTTPPTTTP 1577
QY 402 BKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 461
DB 1578 PPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPSP--PTTTP--PTTP--P 1629
QY 462 TTTKEPAPTTPKEPAPTTPKAP--TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 520
DB 1630 TSITTLPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPPTTTPPTTTPPTTTP 1689
QY 521 KETAPTTPKLTPTTPEKLAPTTPKEPAPTTPBELAPTTPPEPTTTPPEPAPTTPKAAA 580
DB 1690 --SPPTT--TWTTFS--PTTTPSPPTTTTTPSPSTTTP--SPPTTTPPTTTP-- 1735
QY 581 PNTPKAPA-----PTTPKEPAPTTPKEPAPTTP-----KETAPTTP----- 616
DB 1736 --TPSPPTTMTTLPPTTSSPLTTPPSIIPPTSPPTTTPPTTTPCVPCLNWTGMD 1793
QY 617 KG-----618
DB 1794 SGKPNFHKPGGDTLIGDVGCGPWAANISCRATWYDPVPIGQLGQTVVCDVSVGLICKNE 1853
QY 619 -----TAPTTLKEPAPTTPKAPKAPKELAPTTP--TKBPTS 650
DB 1854 DQKPGVIMAFCLNYEINVQCCVCTQPTTM-----TTTTTENPTPTTITTTTPTPTP 1910
QY 651 TTSCKP-----APT--TPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKKA 705
DB 1911 TSTQSPNGLQAPPTPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1969
QY 706 PKELAPTTPKGTPT--TSDKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKPP-----APTPTPT 757
DB 1970 TVLITTTTMTPTPTPTSTKSTTVPTTITTTTATPTTPTTPTTPTTPTTPTTPTTPTTPT 2029
QY 758 PPTTSEVSTPTTK-----EPTTIHKSPDSTPELSAEPKALENSPEKPGPVTTKTPA 812
DB 2030 TPTTGTGPPHTSTAPIALTTNSPPESSTPTSTSTSPSTSTSPSTSTSTSTSTSTSTST 2089
QY 813 ATKPEMTTAKDKTTERDLRTPETTTAAKMTKETATTEKTESKIKTATTOVSTTT 872
DB 2090 STAPSPAPTPTTSGGHTLSPPTTTPSTTTPPTTPTTPTTTP--GSSSAPTPTTPTTSA 2147
QY 873 QDTPFKITT--LKTITLAPKVTTTKTITITTEIMNKPEE-----910


```

QY 234 -----VLAKPTPKAEITTKGPAITTPKE----- 256
Db 5537 ELTESQYFNDRGRCHVMIHNVTDDGQVYSVIARLEPGEARSTAEALVLTKEIKLE 5596
QY 257 -----PPT-----TPKE----- 264
Db 5597 LKPPDIPDSRPIPTMPIRAVPPBEIPVWAPPVLLPTPEEKKPPPKRIEVTKKAVK 5656
QY 265 -----PASTTPKE-----PPTTI----- 278
Db 5657 DAKVAKPKEMTPREBIVKKPPPTTLIPAKAPEIIDVSSKABEVKIMTITRKVEQKE 5716
QY 279 ----- 278
Db 5717 KEAVYEKKQAVHKEKRVFIBSFEPYDELEVETPEFQPYEEDDEDYEBIKVEAKKE 5776
QY 279 ----- 278
Db 5777 VHEEWEDFEGQYEREEGYDEGBEEMEEAYQEREVIQVQKEVVEESHERKVPKVP 5836
QY 279 KGAITTPK----- 286
Db 5837 KKAAPPKVIKKPIEKISRRMEEEKVQVTKVPEVSKKIVPQKPSRTPVQEEVIEVK 5896
QY 287 ----- 295
Db 5897 VPAVHTKKWVISEKMFASHTEEEYSVTVPVEQKEIVTEEKIHVAVSKKVEPPP---K 5952
QY 296 APTTPKEPA-----PTTKKEPAPTTPKEPAPTTPKAPAPTTPKSPA--- 343
Db 5953 VPPELPEKPAPEEAVPPIPKKVEPPAPKVPVEP-----KKVPPEKKVPVPPKKEPAAP 6007
QY 344 --PTTPKKEPA-----TTPKE--PAPTTPKE 365
Db 6008 KVPVPKVPKEKIPVVPVAKKKEAPKVPVEQKGVWTEEKITIVTOREESPPPAVPEI 6067
QY 366 PTEPTTPKE--PAPTTPKEPAPTTPKEPAPTTPKAPAP----- 399
Db 6068 PKKKVPEERKVPVKEEYVPPPKVPA--LPKKVPPEEKVAVVPVAKKAPPPRAEVSXK 6125
QY 400 ----- 399
Db 6126 TVVEEKRFVAEBKLSFAVPQVVEVTRHVSABEWSYSBEEGVSISVYVEREBEEREA 6185
QY 400 -----TTPKEPA----- 406
Db 6186 EVTEYEVMEEPYVVEEKLIHIIKRVAEAPAEVTERQEKIIVLKPAPAKIEEPPPAKV 6245
QY 407 PTPPKE-----PAPTTPKE--PSPTTPKBP--APTTPKSAPTT-----KEPAP----- 446
Db 6246 PEAPKKIIVPEKKVPAPVPPKKEKVPVPPKVPPEEKVPKVPKVPKVIKMEEDLPKAVTEKH 6305
QY 447 -----TTPKSAP--TTPKBSPTTPKE-----PAPT----- 471
Db 6306 MQITQEEKVLNAVITKEAPPKARVPEEPKRAVPEEKVLKPKREBEPKAVTEFRKRV 6365
QY 472 -----PK--EPAP----- 477
Db 6366 KEEKUSIEAPKREPQIKEVTIMEKEKAYLTLEEAVSQREBEEYEEYDYKEFEYE 6425
QY 478 -----TTPKKPAPT--TPKGPAPTTPKE----- 498
Db 6426 PTEYDQVEEYEREVEEHEEYITEPEKIPVKKVPVEBPVTPKAPPAKVLKAVP 6485
QY 499 -----PAPTTPKPAPTAPKEP----- 515
Db 6486 BEKVVPVPPKLLKPPPPKVPPEEKVFBKIHISITKREKQVTEPAKVPMPKRVAV 6545
QY 516 -----APTTPKETAP-----TTPK 529
Db 6546 EKVVPVKEVAPPVVPVPEKLEFEVAFBEEVVTVHVEEYLVBEEBEYIHEEBEFITEE 6605

```

```

QY 530 KLPTTPTEKLAAPTTPKEPAP-----TTPBELAPT----- 558
Db 6606 EVVPEVIVVKV--PEVPRXPVPEEKKVPVPPKKKEAPPAKVPPEVPKKPEKVPVLIPKKEKP 6664
QY 559 ---TPEERTPTTPEEPAPT-----TPKAAAPNTPEKEPAPTTPKEPAPT-----TPKE 602
Db 6665 PPAKVPEVPKVPPEEKVPVPPVPPKKVVEAPPAKVPPEVPKKEVP--EKKVPVPAKKEVAPPA 6723
QY 603 PAPTTPKETAPTTPKGT--APTTLKEPAPTTPK----- 634
Db 6724 KVEVPVKLLIPEEKKPTPVPPKVEAPPKVPKKEVPVFPVALPOEEVILPEEBEIVPEE 6783
QY 635 ----- 634
Db 6784 VLPEEVEVLPEEVEVLPEEVEVLPEEVEIIPPEEVEVPPPEEVEVPEEVEVPEEVLPEVK 6843
QY 635 ---PAP-----KELAP-----TTP 645
Db 6844 PKVPVPAVPEIKKKVTEKKVVIKKEEAPPAKVPPEVPKKVEKRIILPKKEEVLPEVT 6903
QY 646 KEP-----TSTTSKDPAPTTPKGTAPT----- 668
Db 6904 EPEEPEPISEEEIPEBPPPSIEEVEEVAAPRVPEVINKAVPEAPTTPVPPKVEAPPAKVSXK 6963
QY 669 -PKE--PAPTTPKEPAP----- 682
Db 6964 IPEEKVPVPPVQKKEAPPAKVPPEVPKKVVLVPPKKEAVPPAKGRTVLEEKVSVAFRQE 7023
QY 683 ----- 682
Db 7024 VVVKERLELEVVEAEVEEIPPEEVEHVEVEYFEEGFEHVEVEFIKLEOHRVESEHREKV 7083
QY 683 ---TTPKG-----TAPTTL--KEP--APTTPK----- 703
Db 7084 HRVIEVEAEVEVEPEKPAKPPKGPPEISEKIIPPKKPPTKVPVPRKEPPPAKVPPEVKIIV 7143
QY 704 ---PAPKELAPTTTGTSTTSKDPAPTTPKEPAP-----TTPKEPAPTTPK 748
Db 7144 BEKVVPPEPRVPPTKV-----PEVLPPKEVVPPEKKVPVPPPAKKEAPPKVPPEA 7193
QY 749 PAPTTPKE---TPPTTSEVSTPTTKE-----PTTIHKSDESTBELSAEP-- 791
Db 7194 PKEVPEKKVPVPPPKKEV--PPTKVPEVPAKAAVEKKVPEAIPPKPESPPPEVEEPEE 7252
QY 792 TPKALENSPKPEGVPTT-----KTPAA--TKPEMTTTAKDKTTERDLRTPE 836
Db 7253 SPSA--PPKKEVPPVVPVPEVPPEVPEVPEKKVPAAPPKKEV-----TPV 7294
QY 837 TTTAAPKMTKETATTEKTTESKITATTTQVTSITTQDTPPKITTLKTTTLAPKVTTK 896
Db 7295 KVPPEAPK-----EVVPEKKV-----PVPPPKKEVPPPTKVPEV-----PKVAVPE 7334
QY 897 KTIITTT-----EIMNKPEETA----- 912
Db 7335 KKVPEAIPPKPESPPPEVEFEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEEVEE 7394
QY 913 -----KPKORATNSKATTPK-- 927
Db 7395 AVAKKPELPVPPVPEVPPEVPEVPEKKVVLVPPKPEAPPAKVPPEVPEVPEVPEVPEVPEK 7454
QY 928 ---POKRTKAPKPTSTTKP----- 944
Db 7455 PEVPPAKVPPEVPKPEVLEEKPAVVPVPERAESPPPEVVEEPEEIAPEEBEIAPEEBEKP 7514
QY 945 -----KTMPRVRKEKTP----- 957
Db 7515 EEEBPEVPPPAVPEBPKKIIPKVKVPVKKPEAPPAKVPPEVPEVPEVPEVPEVPEVPEV 7574
QY 958 -----TPRKWTS----- 964
Db 7575 PPKEDVKEKIFQLKAIIPKKKVPENPQVPEKVELTPLKVPGEKKVKRLLPERKEPEKEV 7634
QY 965 -----TMEPNPTSRIAEAMLQTTTR-----PNQT 989

```

| | | | | | | | | |
|--|------|---|------|----|------|--|------|-----|
| Db | 7635 | VLKSVLRKRPEEBEPKPKLEKVKKPAVPEPPPKPVEVEVETVTKRERKIPEPTKV | 7694 | Db | 412 | ASKEPLTTLTSLATLSELSTQAE | 471 | |
| Qy | 990 | PNSK-----LVENPKSEDAGAGGETPHMLLRHVFMEVTP----- | 1027 | Qy | 75 | -----FESFER-----QREC-----DCDAQ----- | 89 | |
| Db | 7695 | PEIKPAIPLPAPEPKPE-----AEVKT-----IKPPVPEPEPTPIAAPVTVVVGKKA | 7746 | Db | 472 | VDFPTSGVAGDMETTFENIRAA | 531 | |
| Qy | 1028 | -----DMDYLPVFNQGIINPMLS | 1083 | Qy | 90 | -CKKYDK-----CCPDYESFCA | 123 | |
| Db | 7747 | KAPKEEAAKPKGPIKG----- | 7765 | Db | 532 | TCKNEQDQGRFNMCFNVRVLLC | 590 | |
| Qy | 1084 | SPSP-----ARRITEVWGIPSPIDTV-FTRCNCEGKTPFKDSQY----- | 1132 | Qy | 124 | PVUDEAGSLONGDKPVT--TPD | 173 | |
| Db | 7766 | KTPSPTEAERKLRPGSGEKKPPDEAPPT-----YOLKAVPLKFKVEIKD | 7810 | Db | 591 | TTTASTGSTATASSQTATAGPH | 645 | |
| Qy | 1133 | AGYKPIFKFGGLTGQIIVA-----ALSTAKYKNWPESVYFFKRGSGIQYIYKQ | 1189 | Qy | 174 | NSDTSKETSLT-----VNKE | 203 | |
| Db | 7811 | ILITSEFVGSSAIEFCLVSPSTAIT-----WMKDSNI-----RESP----- | 7849 | Db | 646 | TATTTATSTFTAI PSSSLGTTW | 705 | |
| Qy | 1190 | GRPALNPVYGEWTVRRRRF-----ERAIGSQTHTIQYSPA-----RLAYQDK | 1237 | Qy | 204 | STDGKEKTTSAKETOSIEKTS | 238 | |
| Db | 7850 | -----KHRETIADGKDKL-----HIIDVQLSDAGEYTCVLRGNKEK | 7886 | Db | 706 | GATGSVATPS-----STPGTA | 755 | |
| Qy | 1238 | -----GVLHNEVKV-----SILWRGLPNVVTSAISLPN | 1265 | Qy | 239 | -----TPKAEITTKG--BAL | 258 | |
| Db | 7887 | TSTAKLVVEELPVRFKLTLEEEVTVVKGQPLVLSCELNKERDVWR----- | 7932 | Db | 756 | TTPTTRGSTVTPSSIPGTHT | 811 | |
| Qy | 1266 | IRKPDGYDYAFSKQYVNI | 1296 | Qy | 259 | PTT-----PKEPASTTPKEPT | 295 | |
| Db | 7933 | -----KDGKIVVEKGRIVPGVIGLMRALT | 7957 | Db | 812 | ATTIATG | 871 | |
| RESULT 4 | | | | Qy | 296 | -----APTTP-----KEPAP | 317 | |
| T45025 | | | | Db | 872 | THGRSLSPSPHTVCTAMTSAT | 931 | |
| C:Species: Homo sapiens (man) | | | | Qy | 318 | EPAPTTKEPAPT | 346 | |
| C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000 | | | | Db | 932 | PSRITTESPPSGTTTTPGHT | 991 | |
| C:Accession: T45025 | | | | Qy | 347 | ----- | 346 | |
| R:Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A. | | | | Db | 992 | QCAMSEWLDYVPMPOGPGSGD | 1051 | |
| A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat | | | | Qy | 347 | -----PKKPAPTTKEPAPT----- | 361 | |
| A:Reference number: 222899; MUID:97166151; PMID:9013550 | | | | Db | 1052 | VECSLDFGLVCRNEQVGKFMCF | 1111 | |
| A:Accession: T45025 | | | | Qy | 362 | -----TPKEPTTTKEPAP-----TTKEPAPT-----TPKEPAPTAPKKA | 398 | |
| A:Status: preliminary; translated from GB/EMBL/DBJ | | | | Db | 1112 | TELTATTTESTGSTATPTSLRT | 1171 | |
| A:Molecule type: DNA | | | | Qy | 399 | -----PTTPK-----EPAP-----TTPEPAPT | 423 | |
| A:Residues: 1-3570 <DES> | | | | Db | 1172 | LRSTATTATATSVTPFPSSSLG | 1231 | |
| A:Cross-references: EMBL:272496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503 | | | | Qy | 424 | -----TTPEPAPT | 456 | |
| A:Experimental source: placenta | | | | Db | 1232 | TTTGATGSVATPSSTPGTATTK | 1291 | |
| C:Genetics: | | | | Qy | 457 | ----- | -K | 457 |
| A:Gene: MUC5B | | | | Db | 1292 | STVTPSSIPGTHATVLT | 1351 | |
| Query Match 14.5%; Score 1026.6; DB 2; Length 3570; | | | | Qy | 458 | EPSPTTTTKEPAP-----TTPEPAPT-----TPKKPAPTTPKEPAPTTP----- | 496 | |
| Best Local Similarity 13.5%; Pred. No. 2.2e-13; | | | | Db | 1352 | NPSSTPGTRPIPPVLT | 1411 | |
| Matches 472; Conservative 123; Mismatches 531; Indels 2374; Gaps 106; | | | | Qy | 497 | -----KEPAPTTTKPAPT----- | 514 | |
| Qy | 23 | SSQDLSSCAGRCG-----EGYS-----RDATC----- | 44 | Db | 1412 | HTVKTAMTSATSGTLGTHTEP | 1471 | |
| Db | 172 | SAPGTTTCQPRCQWTEWDEDPKSEQLGGDVESYDKIRAGGHLCOQPKDIEQAESFP | 231 | Qy | 515 | PAPTTP-----KETAPTTPK-----LPTTPEKLAPTTP----- | 544 | |
| Qy | 45 | -----NCD-----YNCQHYMECCPD----- | 59 | Db | 1472 | PGTTTPOGHTATSR | 1530 | |
| Db | 232 | NWTLAQVGQKVCVDVHFGLVCRNWEQGVFKMCNRYRVLCCSDHCRGRTATPPPTTE | 291 | Qy | 60 | ----- | 74 | |
| Qy | 60 | ----- | 59 | Db | 292 | LETATTTTQALFSTPQPTSSPGL | 351 | |
| Db | 292 | LETATTTTQALFSTPQPTSSPGLTRAPPASTTAVPTLSEGLTS | 351 | Qy | 60 | ----- | 59 | |
| Db | 352 | QSAGSTETVPGVATSLTPTRSALPGTTGSLGTWRSPQPTLAPTWTATSPARTGTAST | 411 | Qy | 60 | -----FKRVCVTAELS-----CKGR----- | 74 | |

```

QY 545 -----EXP----- 547
Db 1531 SYMPGPGGDDFYSNIRAGAVCEQLGLECRAQPGVPLRELQGVVCESLDFGLV 1590
QY 549 -----APTPEELAPITPEE----- 562
Db 1591 CRNREOVGKFCNFVEIRVFCNCGHCPSTPATSSSTATPSPTGTTWLLTQTTAAIT 1650
QY 563 -----PTPTTPEPAP-----TTP-----KAAAPNTPK----- 585
Db 1651 ATTGSTALPSSPTGAPPPKVLTSQATPTATSSKATSSSPRTATLPLVLTSTATKSTA 1710
QY 586 -----EPAPTT-----PKEP-----APTTPKEPAPT 606
Db 1711 TSFTPIPSLTLTGTSQNRPPHPMATMSTIHPSSTPETHTMSTVLTKAITTRATSSMS 1770
QY 607 TPKEPAPT-----PKGTAPT-----LKEPAPT----- 630
Db 1771 TPSSTPGTTWILTELTAAALPHGTPSSPTGTTWILTEPSTTATVTVPTGSTATAS 1830
QY 631 -----TPKK-----PAKELAPT----- 643
Db 1831 STRATAGLKVLTSTATPTTVISSRATPSSSPGTATAPALRSTATPTATSVTAIPSSS 1890
QY 644 -----T-----T 644
Db 1891 LGTAWRLSQTTPTATMSTATPSSTPETVHTSTVLTTATTTRGVSATPSSTPGTAHT 1950
QY 645 TKEPSTT-----SDKPAFT----- 660
Db 1951 TKVPTTTTGTATPSSSPGTALTPVWMSITTTTTRGSTVTPSSIPGTHTAIVLTT 2010
QY 661 -----PKGTAPTTPKEPAPT----- 676
Db 2011 TTVVATGSMATPSSSTQTSPTLTTATTATATGSTTNPSSTPGTTPIPVLLTTATT 2070
QY 677 -----PKEP-----APT----- 683
Db 2071 PAATSSVTPSALGTHTPPVNTTATTHGRSLPSSPHVPTAWTSATSGILGTHIT 2130
QY 684 -----TPKGAPT-----LKEPAP-----TTPKKPAKELAPTTTKGPTSTTSKPA 726
Db 2131 EPSTGTSHTPAATGTTQSTPALSSPHSSRTTESPPSP-----GTTTPGHTRGSRRTA 2186
QY 727 PTTKETAPT-----PKEP----- 741
Db 2187 TATPSKRTSTLLPSSPTSAPITTVVTGCBPQCAWSEWLDYSYMPMPGSGGDDFYSGNI 2246
QY 742 ----- 741
Db 2247 RAAGGAVCEQLGLECRAQPGVPLRELQGVVCESLDFGLVCRNREOVGKFCNFVEI 2306
QY 742 -----APTTKKPAPTTPPTPT----- 761
Db 2307 RVFCNCGHCPSTPATSSSTATPSPTGTTWILTKLTATTTESTGSTATPSSTQGPAG 2366
QY 762 -----SEVSTPTTT----- 770
Db 2367 TPHVSTATTPTVTSSKATPSSPCTATAPALRSTATPTATPTATPSSSLGTTWTRL 2426
QY 771 -----KEPTTIHKS----- 779
Db 2427 SQTTPMATMSTATPSSTPETVHTSTVLTTATTGATGVSATPSSPTGTAHTTKVPTTT 2486
QY 780 -----PDEST-----PELSAETPKALEN----- 798
Db 2487 TTGTVTPSSSPGTARTPPVMSITTTTPTSGSTVTPSSIPGTHTPVLTPTVLTQPVATG 2546
QY 780 -----PDEST-----PELSAETPKALEN----- 798
Db 2547 SMATPSSSTQTSPTLTTATTITATGSTTNPSSTPGTTPIPELTTATTATPAATSS 2606
QY 799 -----SPKEP----- 803
Db 2607 VTPSSALGTHTPPVNTTATTHGRSLPSSPHVPTAWTSATSGTLGTHITHEPSTGTS 2666
QY 804 -----GVPTTKTAAATKPEMT-----TTAKDKTTER----- 829
Db 2667 HIPAATTGTTTSTFALSSPHSSRTTESPPSPGTTTGHHTTATSKTTATATPSKTRTST 2726
QY 830 ----- 829
Db 2727 LLPSQTSAPITTVVTIGCEPOCAWSEWLDYSYMPMPGSGGDDFYSNIRAGAVCEP 2786
QY 830 ----- 829
Db 2787 LGLECRATAQPGVPLGELGVVCESLDFGLVCRNREOVGKFCNFVEIRVFCNCGHCP 2846
QY 830 -----DLRTTPETT-----TAAPKMTKETA-----TTTE 853
Db 2847 STPATSSSTAMPSTPGTTWILTELTATTASTGSTATPSSTPGTAPPPKVLTSPTATP 2906
QY 854 KTTESKI-----TATTQV-TSTTQDT-----TPFKITTLKTTTLAPKVTTTK-KTIT 900
Db 2907 TATSSKATSSSPRTATLPLVLTSTATKSTATSVTPIPSSILGTTGLPEQTTTPVATMS 2966
QY 901 TTEIMNKPEE-----TAKPKDRATNSKAT-----TPK 927
Db 2967 TIHPSSTPETHTMSTVLTKATTRATSTSTSPSTPGTTWILTELTAAATTTAGTGTAT 3026
QY 928 POK----- 930
Db 3027 PSSPTGTTWILTELTATTASTGSTATLSTSTPGTTWILTEPSTTATVATPPGSTATAS 3086
QY 931 -----PTKAPKPTSTKPKXT---MPPVRKPKTTPTPRKM----- 963
Db 3087 STOATAGTPHVSTTATPTVTSKATPSSSPGTATAPALRSTATPTATPSFTAISSSL 3146
QY 964 -----STMPELNPTSRAEAMLOTT-----TRNQTNPNSKLV 995
Db 3147 GTTWRLSQTTPTATMSTATPSSTPETVHTSTVLTTATTGATGVSATPSSTPGTAHT 3206
QY 996 EWNPKSEDAG-GAEGETHPMLLRPHVEM-----PEVTPDMDYLR 1034
Db 3207 TKVPTTTTGTATPSSSPGTALTPVWMSITTTTPTTPTTSGSTVTPSSIPGTHHTAR 3266
QY 1035 VPN-----QGIINPMLSDETNICNGKPDVGLTLLEN-----GTLVAFR 1073
Db 3267 VLTSTTTTATGSMATPSSSTQ---SGTPPSLTTATTITATGTTNPSSTPGT----- 3318
QY 1074 GHYFWMLSPSP-----PSPA 1089
Db 3319 -----TPIPPVLTSMATTPAATSSKATSSSPRTATLPLVLTSTATKSTATSFTPTPS 3371
QY 1090 RRITEVMGIP-----SPIDTVEFRNCE-----GK 1114
Db 3372 STLWTTWTVPAQTTPMSTMTSIHTSSPTETHSTVLTTTATWTRATNSTATPSSILGT 3431
QY 1115 TFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQOIVAAL-----STAKYKNWP 1163
Db 3432 T-----RIITELTTTATTAAATGSTATLSSTP 3458
QY 1164 ESVYFFXEGGSIQOYIYKQEPVQKCPGRPALNVPVVGEMTOVRRRPERAIGESQHTI 1223
Db 3459 GTTWI-----LTE-----PSTIATV 3473
QY 1224 RIOYSPARLAVQDKGVHLNKKVSIILWRGLPNVVTSAISLNPINRKPDPGYDYAFSKOQY 1283
Db 3474 MVPTGSTATTSSILGTAHT-----PKVVITAMATMP----- 3503
QY 1284 NIDVPSRTARAITRSGQTL 1303
Db 3504 -----TATASTVPSSSTV 3516

```

RESULT 5

T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotcov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing-in of a proline-rich exon converts alphanac into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:gi1666688; PID:gi1666689; PID:AAB187
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphanac into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 14.5%; Score 1021.4; DB 2; Length 2187;
Best Local Similarity 20.4%; Pred. No. 9.6e-14; Indels 896; Gaps 81;
Matches 392; Conservative 159; Mismatches 479; Indels 896; Gaps 81;
4 KTLPIYLL-----LLSFVVIQVSSQDLSCAGRCGEYSR 40
Db 479 KNLPIALVNVGAPVSPAQAGLPRKDTTLQPLALPKESPSQASSL-----EVLSE 533
QY 41 D-----ATCNCYCNCHMECCDFKVCITAEISCKRCFCSEFE 79
Db 534 DTVTKTGGPAPVVRPAIGVAT-----TTL----- 561
QY 80 RGRECDCAQCKYDKC-----C 97
Db 562 ---RADSPAVIRADSCVSPNTVQPLKESVTDPMAPRTAKNTAPSTTSLVPLASGCG 618
QY 98 PDYESFCRAVDKNNRT-----KKKTPPKPV-----VDEAGSGLDNGDF 138
Db 619 PVASSMALSPQNASVETALASPEIKSVPPDPLAEISFSPNARKVDVSHMESGSS 678
QY 139 KVTTPDSTTQH----- 150
Db 679 RQGHDPASVTAKGTVVCLADSLDTSVASKGSLSGASSPLYPLEVSLPEAGLAVQGP 738
QY 151 ---NKVSTSPKITTAKPINRPSLPNSDTSKTSITVKNKETTVEIK----- 195
Db 739 KGSNLKSLTPPSSKGAPV-PSTGAP---SPKGAPIVPTESSISSKQVPAEILPSPQK 793
QY 196 ---TTTNKOTSDGKEKTSKAKETQSIKTS----- 225
Db 794 TPEVTASRLISAVQPKVDPMDSVDVTPSPKTSATAVPKDTS-----TSLKSVPAVTSLS 851
QY 226 ---KDLAPTSKVLAP-----TPKAEITTKG 248
Db 852 PPKAPVAPSNEATIVTEIPTSLKNALAAATPKETLATSIPKVTSPSPKTPKSVSLKGA 911
QY 249 PALTT-----PK-----PTPTTPKEPASTT-----PKEPTT 275
Db 912 PAMTSKATEIAASKDVSPQFKEVPLLOHVPTTSPPKSPVSDTILSGALTSPPKGP-P 970
QY 276 TTIKSAPTPKE-PAPTTTKSAPTPK-----EPAPTTKEPAPTTKEPAPTTTK 325
Db 971 ATLAEPTYPKSPKPAASKTPTATPSPBGVAVPLEIFPCGKAPKTAAPKESATSS 1030
QY 326 EPAP-----TTTKSAPTPKEPAPTTPKK-PAPTPKE-- 357
Db 1031 KRAPKTVANSKEIPSGVAVPLEISLPKETSKA--TGEKSASSPKSPKTAGKETP 1088
QY 358 -----PAPTPKEPTP-----TTPK----- 373

Db 1089 PGGVAVPPEISLPPKETPONATPNESLAASSQKRSPKTSVPKTPPGVGTAMPLEIPSA 1148
QY 374 --PAPTTKEP-----APTTKEPAPTAPKKPAPTTP----- 402
Db 1149 PQKAPKTAVPKQIPTPEDAVTILAGSPUSPKASKASTAAKPEAPATPSVGVIAVSGEISPS 1208
QY 403 -----KEPAPTTKEPAPTTTKE-----PSPTTPKE---P 429
Db 1209 PKKTSKTAAPKENSATLPPKRSKPTAAKPE--TPATSSGVTAVPSEISPSPTTASKGVP 1267
QY 430 APPTTKSAPTTTKBPAPTTTKSAPTT--PREPS--PTTKBPAPTTTPKEPAPTTPKKPAP 485
Db 1268 VTLTPKGAPNALAE-SPASPKKVPKTAAPETSTTPSPQIKPVAGPKAASATPSKKTTP 1326
QY 486 TT-----PKEAPPTTPKEPAPTTTKBPAPTPAKE-PAPTTKEPAPTTPKKLTPTTP 536
Db 1327 KTAVPKETSAPSEGVAVPLEIPSPRKAPKTAAPKETPAPS--PEGATTAPVQI-PPSP 1383
QY 537 EKLAPTTPEKAPTTPEELAPTTPEEPTPTTPEEPAPT--TPKAAAP----- 581
Db 1384 RGSKKAGSKETPTTP-----SPGVTAAPLEIPISKKTSKOMASPKETLVTSPSKKLS 1437
QY 582 -----NTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEPAPTTPKKP 635
Db 1438 QTVGKETSELEGATAVPLEIPSPHKKAPKTVDPKQVPLTPSPKDPAPTTLAE-SPSSPKK- 1495
QY 636 APKELAP---TTTKEPSTTSDDKAPTTPKGTAPTTPKEPAPTTTPKEPAPTTTPKP 682
Db 1496 APKTAAPSPSERVTTVPP-----EKDA-TPQKASGTTASKVPVPAETQEVAVSSRRTPTVP 1549
QY 683 TTPKGTAPTTLKEPAPTTPKKPAPKELAPTTTKGPTSTTSOKPAPTTPKETAPTTPKEPA 742
Db 1550 AVPPVKNSSHKKTSKTIELKEAPATLPPSPKSPKIPSSKK---APRTSAP----- 1598
QY 743 PTPPKKAPAPTTPETPPPTTSEVSTPTTKEPTTIHKSDESTPE--LSAETPKALENSP 800
Db 1599 -----KEFPASPSIKPVTSLAQT-----APPSLOKAPSTTIPKENLAA---PAVLVSS 1645
QY 801 KEPGVPTTKTAAATKPEMTTAKDKTTERDLTTP-----ETTTA-----APK 843
Db 1646 KSPAAP-ARASASLSP---ATAAPQTAPKEATTIPSCKKAATETPIETSTAPSLEGAPK 1701
QY 844 MKTETATTTEKTESKITATTQTSTTTTQDTPPKITTL---KTTTLAPKVTTTTKTI 899
Db 1702 ETSEISV-----SKVLMSSPPKASSSKRASTLPAITLPSLKEASVLSPTATSGK-- 1752
QY 900 TTEIMNKEETAPKDRATNSKATTPKPKQKTKAPK----- 937
Db 1753 -----DSHISPVSDACSTGTTT--PQASEKLPSKKGPTAFTMLAAPAPESALAIT 1801
QY 938 -----PTSTKKPKTMR--VRPKTTTTPKMTSTWPELNPTSR 974
Db 1802 APIQKSPGANSNASPKCPDPSKKOTKGLPSAVALAPQTVPEK-----DTSK 1851
QY 975 TAEAMLQTTTRENQ----- 988
Db 1852 ALETLLVSPAKSDCLHSPKPGVGSQVATPLAAFTSDKVPPEAVSASVAPKPAASLTL 1911
QY 989 ---TPNSKLVENVPKSEDAGGAEGETHMLLRPHVEMVEVTPDMDYLPVFNQGIINPML 1046
Db 1912 APSVPAPLPPKQPLLESAPGSVLESPSKL-----PVPAAEDEL-----PLI 1953
QY 1047 SDETNICNGKPGVDGLTLRNGTLVAFRGHYFWMLSPFSP-----PSPARRITEVWGLP 1099
Db 1954 PPEA-VSGE-----PFQDILNMAPKPA-----GTP 1980
QY 1100 SPIDVTFRNCNCEKGTFFFKDSQYWRFTNDIKDAGYKPIFKGFG----- 1145
Db 1981 APAPSA-----KQVLLKNNKSGSTESDESVPBL 2010
QY 1146 -----LTGOIWAALSTAKYKNWESVYFFKRGSGSIQOYIYKQEPVQKCPGRPAL 1195
Db 2011 BEQDSTQTATQAAQAAAAE-----IDEEPVSQAKQSR--- 2043


```

QY 380 -----EPAP----- 383
Db 1583 EDINECVAEKAPCSLNANVMNCTFSCSKQGYRGDFMCTDINECDERHPCHPHAECT 1642
QY 384 -----TTPKE----- 388
Db 1643 NLEGSFKCECHSGFEGDGKIKCTNPLERSCEBDEVEKFCGRVDHVSLSVRIYNGSLSSVCE 1702
QY 389 -----PA----- 390
Db 1703 CEPGRFEKESNCVDIDECEBSRNNCDPASAVCNTEGYSRCECARGYEGEGVCTDID 1762
QY 391 -----PTAPK----- 395
Db 1763 ECDRGAGCDMSAMCINRMSCGCKMAGYTGDCATCIKEEPKSKDTACTDEWSLCE 1822
QY 396 -----KPAPTTKPEAP----- 407
Db 1823 LEKKQCTVDEEVPQCGACLPFGHHPINGTCQSLQISGLCAQKNCNKHAECIDIHPSHF 1882
QY 408 -----TTP 410
Db 1883 CSCPDGFIGDMICDDVDECNMAGMCDENTKENTIGSFNCVCLGFKVKVDEKCVVDEK 1942
QY 411 KEP-----APTITKEPSPTTKEPAPTITTKSAPTITKEPAPTITTKSAPTITKE 458
Db 1943 KQPNREKIEIDENSSSSNGQEPKTKGIVSSTSATSSSTTAEBHTVTSISSTSTKD 2002
QY 459 PSPTITKEPAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPAPTTPKKPAPTAPKEPAPT 518
Db 2003 MTSSKSPENVMTSESFEVSTSSSKSTTASSETTVSSSTPSSS-----SSAPLTSSPA 2055
QY 519 TKETAPTTPKKLTPTTPE-----KLAPTPE-----KPAPTITPEELAPTTPPEP 563
Db 2056 TTTEV-----ITESSVKSTTPKESSSEITVKLSSKSPETVSSSVKSSPSTPS-----TTSQSV 2109
QY 564 TPTTPEEAPTTPKAAAPTTPKEP-----APTTPKEPAPTTPKEPAPTTPKET----- 611
Db 2110 TSTVPETSKSTVLSSEAPTSTSPTEVHTSSSETKPSLSASSITGDNSTTPSTSSLASVK 2169
QY 612 APTTPKGT-----APTILKEPAPTTPKKAPKELAPT-----TTKEP----- 648
Db 2170 STSAPEGTSASVAPVKLSLSPDV-SQPSTKTFDATESSTVQASSETSGTSVKSTSEPS 2228
QY 649 -----TSTTSDKAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKG----- 687
Db 2229 HVTKLSITSSNFSSSVPTSPKSTPTVPPEST-EQPTSTTPSGQSLTPMNSNSELVTTSE 2286
QY 688 -----TAPTTLKEPAPT-TPKKAPKEL-----APTITKGT-----ST 720
Db 2287 PHVLSSSLSPDYSQSTTENNLSSESTVETPKTSSSEVLSNSEPSTTEAPTTLSPDLST 2346
QY 721 T-----SKP-----APTTP----- 730
Db 2347 TTNNLSQSTVSTEDRSEISSENSEKPTSAPELVTSSVTHVASSPDVPTESSEPDLLTG 2406
QY 731 -----KETAPTTP-----KEPAPTTPKKP-----APTPE----- 755
Db 2407 SSTEMIPEASSKQTLSTPTPTTTTASEEPTKSTMSPLSTTSNVLSESTTPSSSSKS 2466
QY 756 -----TTPPTTSEVSTPTTKEPTT----- 775
Db 2467 PVSSSTEGISVVTSTEFKSVPESTISSVLEEDLTKTTPSPILTEETTTASETSEPLETDSL 2526
QY 776 -----LHK--SPDESTPELSAEPPTKALENSP-KEPG-----VPTT----- 808
Db 2527 TVSVRIHELTTSENVPKSESTTTSSSKSPSQEPAGILTSTVVVPTSSVSLITASIE 2586
QY 809 -----KTPAATKPEM-----TTTAKDKTTERDLRTPPTTTAAPKMTKETA 849
Db 2587 AITSNTPKQGRTPITTPGKSLVKSTTSPSTVTSSEPSSESTKRITVSTVSTTPTTEBT 2646

```

```

QY 850 TT-----TEKTTESKITATTQVTS-----TTTQDTTPF 878
Db 2647 TSESLILITAAPSKPTTESTTESSEAPTTPAKTSETKPNVSVSTSKSTENVETSSQSGSL 2706
QY 879 KITTLKTT-----TLAPKVTTTKTITTT----- 902
Db 2707 ESSIMSTSEPEPTNAPAVTVSSEASTTLEENSTSTSPSSSEASVKLSLFPESITSEA 2766
QY 903 -----EIMNKPEE----- 910
Db 2767 VTVSSRAPAEITMSSESHREISTVSSEPSPEPLSTTVSPNVVTVASSISEEPILSVT 2826
QY 911 -----TAKPKD-----RATNSKA-----TTPKPKQ 930
Db 2827 SSSTPRVLLTGTGPDLLIVSTVPSHGNNRRQNTASSVPSNSTSPILPSESLLTTPQPPP 2886
QY 931 PTKAPKPTSTTKPKMVRVRKPK-----TTPETP-----RKWTSTMPELNP 971
Db 2887 TTTTAKPATTSKGRGPPSIQPPAEMFTTPAPPPNSNGYGEETNOEEQVTS----- 2940
QY 972 TSRIASAMLQTTT----- 984
Db 2941 TTTEAPSLCSTVTCCHSLATCEQSTGVCIQDGFIDGTTACSKKSTADCLISLCLADKA 3000
QY 985 -RPNQTPNSKLVEVNPKSEADAGAGETPHMLLRPHVFMPEVTPDMDYLPVFPVQGIIN 1043
Db 3001 KCDNSTRSCEC-----DAG-----YI-----GDGVCS 3023
QY 1044 PMLS-----ETNICNGKPV-----DGLTLR-----NGT 1068
Db 3024 PHQDCVLRDLNLCSPAVQNRRCCQLPGFTGDKVKCSIHERASNCQCDANAHCVGT 3083
QY 1069 LVAFRGHYFMLSPFPSPARRITEVWG-----IPSPIDTV-FI-----R 1108
Db 3084 TCKCNPGYF-----GNGLCCVPDPLDCVHTFGICHFNVCNSESQ 3124
QY 1109 CNCEGKTFPFKSOYWRFTNDIKDAGYKPIFKGFG-----LTQIIVAAALSTA 1157
Db 3125 CQCS-----SGFSNGVSCFPPOKSCRTDKSVCA----- 3152
QY 1158 KYKNWPESVYVFPKGGSIQYIYKQBPVOKCP-----GRRPA-----LN 1196
Db 3153 -----KNAICLPTGSCICRHGFKGDPFYKCSLSVSKEPANQODLSDVSCVTPCDSSTQ 3206
QY 1197 YPVYGMTQVRRRRFER--AIGPSQT----- 1220
Db 3207 LCISGEC--ICKSGFRNSTLSSSETCADIDECAEKSHKCDRVATCRNTFGSHVCTCPDG 3264
QY 1221 -----HTIRIQSPARLAYQDKGVLHNEVKVSLWR--GL----- 1253
Db 3265 HVGDGITCVPHV-----NOGKLSVYCEADGMTVLGLNETADPEGI 3305
QY 1254 -----PNVVTSAISLPNIRKP----- 1269
Db 3306 FVKGGAENPYCKSPSSLLNSRKPVFKVFOHCDVQLLDNHTMASTVVVQKHAMFLTNK 3365
QY 1270 -DGYDY-----YAFSKDOVYNID----- 1286
Db 3366 ADSYDLRCQYPIGSRVESHVNVSELAITSTLTTEKNSTLAPICRLSVSNDQHSISSAMV 3425
QY 1287 -----VPS-----RTARAITRSGQ 1301
Db 3426 GDTLKALEVTPSENFNGILPRNCFAINTESGE 3457

```

RESULT 10

T29757

Protein UNC-89 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999

C:Accession: T29757

R:Du, Z.; Le, T.T.; Wilson, R.

submitted to the EMBL Data Library, May 1997

A;Description: The sequence of C. elegans cosmid C09D1.
 A;Reference number: Z20679
 A;Accession: T29757
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-6642 <DUG>
 A;Cross-references: EMBL:AF003131; PIDN:AA854132.1; GSPDB:GN00019; CBSP:unc-89
 A;Experimental source: strain Bristol N2; clone C09D1
 C;Genetics:
 A;Gene: CESP:unc-89
 A;Map position: 1
 A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
 /3; 5917/1; 6021/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1
 Query Match 13.4%; Score 945; DB 2; Length 6642;
 Best Local Similarity 12.2%; Pred. No. 4.9e-11;
 Matches 480; Conservative 190; Mismatches 470; Indels 2780; Gaps 123;
 QY 34 CCEGYSRDATCNG-----DYNCOH----- 52
 Db 659 CKEGVIDFCEBVEGWPELWLVDDQPLRSHDFRLQYDGGTAKLEIRDAQPDGTGY 718
 QY 53 -----YMECCPD-----FKAVCTAE-- 67
 Db 719 TVKIQNEFGSIESKAEFLVQADFDKNHVAPEFOATIEYVEC--DEGEVRFKSVITGDPN 776
 QY 68 -----LSCKGRCPESFERGECDC 86
 Db 777 PEIIFWINGKLSEKVKFISEDGICILTIKDVTRHFDGMVTCOS-----NRLGSASC 831
 QY 87 DAQCKYDKCCPYESFCAEVKDNKNRTKKTKPPPVUDE----- 128
 Db 832 D-----GRUKVRVPAPPTFNKELEBDKTVQEKSTVVVFVD 866
 QY 129 -----AGSLONGD-----FKVTTPTDTSTQHN----- 151
 Db 867 VSGWPEPTLTFTLCGKELNGBEGVEIVGHDFYRISIPNTSMDKHGEIVAKAQNEHGT 926
 QY 152 ----- 151
 Db 927 AESRRLTVQEBEERSAFTFLKIDBQTVKTGEFAVFETTVRGNPNPVTWFINGHKM 986
 QY 152 ----- 151
 Db 987 DOGSPGVKIEAHNDHKLITDSQAYAGTVLCRAENAVGRPETKARLVVLAPKQKPPKF 1046
 QY 152 -----KVSTSPKLTIA----- 162
 Db 1047 VEILVDKTEVDNTVVFEVRVGEFKPTVTWYLGEBELKQSDRVEIREFDGSIKISIKNI 1106
 QY 163 -----KPINRPSLPP----- 173
 Db 1107 KIEDAGEIRAVATNSEGSDETKAKLTQKKFPAPFEDLRVSLTVKEGSAVFAHAFGI 1166
 QY 174 ----- 173
 Db 1167 PLPTYEWSVNGKVRQDQEGARVTRDESTVDGASILIDITATYYSEVNNHLTISVAENTL 1226
 QY 174 -----NSDTSKE-----TSLTVNKE 188
 Db 1227 GAETGAQLTTEPKKESVWVEKQDLSSEVQKEIAQQVKEASPEATTITMETSLTSTKT 1286
 QY 189 TTVEETKATTT----- 199
 Db 1287 TTMSTTEVTSVGGVTVEKSESASATTVIGGGSGVTEGSIKSVSVKIEVVVKTDSQTDV 1346
 QY 200 -----NKQSTDGKKEKITSKAKETOSIEKTSKADLAPTS 232
 Db 1347 REGTPKRVSAEBELPKEVIDSDRKKKSPDPKKEK-----SPEKTEEPASPTK 1398
 QY 233 KVLAKPTPKAETTTKGPALTTTPKEPTTP-----PTTIKS 280
 Db 1399 KT-----GEEVGPKEKSPASPTKSKSPAAEVRVKSPTTKKEKSFSSPTKKEK 1445

QY 281 APTTP-----KBPAPTTTKSAPTTPKEPAPTTTKBPAPTP-----KEBPAPTTT 324
 Db 1446 SPSSPTTKTGEVKEKSP-----PKSP---TKKEKSPKEDVKSPVKKEKSPDAT 1493
 QY 325 -----KEBPAPTTTKSAPTTP----- 339
 Db 1494 NIVEVSSETTIETKTETMTMETHESESRTSVKKEKTPKVDKPKSKTKDKSPEKSI 1553
 QY 340 -----KEBPAPTTTKSAPTTPKEPAPTTTPKEPTTPKE-PAPTTKEPAP----- 383
 Db 1554 TBEIKSPVKKEKSPKVEKBPASPTKKEKSPKBPASPTKKEKSPVKKEKSPKSVV 1613
 QY 384 ---TTPKEPAP-TAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPT 439
 Db 1614 ELKSPKESPEKADDPKSPKKE-----KSPEKSAATEDVKSPKTKKEKSPKVEKPT 1667
 QY 440 TTKBPAPTTTKSAPT-----TPKEPSPTTTKE-PAPTTTPKEPAP-----TTP 480
 Db 1668 -----SPTKKESSPTKKTDEVKSPTKKEKSPQTVBEKSPASPTKKEKSPKVEKSP 1722
 QY 481 KKPAPTTTKE-PAPTTTPKEPAP-----TTTTK-----PAPTAPKEPAPTTTPKETAPT 527
 Db 1723 KKSPEKABEKKSPKTKKEKSPKGAEEVKSPKTKKEKSPKGAEEKSPKTKKEKSPV- 1781
 QY 528 PKKL-----TPTTPEKLAPTTTPKEPAPTTTPBELAP--TTPPEPTTTPPEPAPTTPKAA 579
 Db 1782 ---KMADEVKSPKTKKEKSPKVEKFPASPTKKEKTPKSAABELKSPKTKKEKSPSPTKK 1839
 QY 580 APNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTLKEPAPTTTPKPKAP- 638
 Db 1840 TGDSEKESPEKPEE---KPKSP---TPKSPGSPK-----KKSKSPEAKPPAPKL 1887
 QY 639 -----ELA----- 641
 Db 1888 TRDKLQTNKIDLAHFVVVVEHATECKWFLDGEKITTAGVTVSKDDQEPFRCSDITM 1947
 QY 642 -----PTTTKEPTST----- 651
 Db 1948 FSGSTVSVASNAAGSVETKELKVLTPKETKKEPFTDKLRDWEVTKGDTVQMDVIALH 2007
 QY 652 -----TSDKAPATT 660
 Db 2008 SPLKYQWQNGILLDGKNGVTIKNEENKSSLIIPNAQDSGKITVBAENSVGSSSAQLT 2067
 QY 661 --PKGTAPTTTPKEPAPTTTKE-----PAPT----- 683
 Db 2068 VNPSTTPIVVDGPKSVTIKETETAEFKATISGFPAPTVKWTINEKIVBSRITTIKTE 2127
 QY 684 -----TPK----- 686
 Db 2128 DVYTLKISNAKIBQGTGVKVTQAQNSAGDSKQADLKVEPNVAKPKFSQLTDKVADEGEP 2187
 QY 687 -----GTAPTTLKEPAP----- 698
 Db 2188 LRWNLELDGSPGTEVSWLLNGOPLTKSDTVQVVVDHGDGTYHVTIAEAKPEMSGTLTAKA 2247
 QY 699 -----TTPK-----KPAPK----- 707
 Db 2248 KNAAGECETSAKVTVNGGNKKPEFVQAPQNHETTESVKESAIVTGKPMNVWYLNKK 2307
 QY 708 ----- 710
 Db 2308 KLIQSEEVKVVYHETGKTSIRIQKPLMEHNGTIRVEAENVSGKVQATAQLKVDKKEVTP 2367
 QY 711 PTTT-----KGPTSSTSKP-----A 726
 Db 2368 KFTTMDRQVKEGEDVKFTANVEGPEPSPVAVMTNNGEPVSKPHNIITVTDKGHTHEIS 2427
 QY 727 PTTTKEP-----APTTK----- 739
 Db 2428 AVTEQAGELSCBATNPVGSKKRDVQLAVKKVGADPTFAKNLEDRLITEGELTLMDAKLN 2487

QY 740 ----- 739
 Db 2488 IVKPKPTIWLKGVETSDGHYKIVEEDGSLKLSILOLKLKEDKGRITIKAESFGVAE 2547
 QY 740 ----- 745
 Db 2548 CSASLGVVGRPMKPAFQSDIAPINLTGDTLEKLLITGDTFFVKWVIGTQVCATE 2607
 QY 746 ----- 756
 Db 2608 DTEISANGVYTMKIHGVADMTGKIKCVAYNKAAGEVSTGCPKLVKVAIPVFEFETSLCDA 2667
 QY 757 ----- 775
 Db 2668 TCREGDTLKLAVLLGEPEPVVSWYNGKLBESQNIKHSEKGTVTYVITKIDTCYSGQ 2727
 QY 776 ----- 778
 Db 2728 WVCEAINEYKATSEATLVLPRGEPDFLEWLSNVRARTGTQVHKVWFTGDPKPSLTW 2787
 QY 779 ----- 781
 Db 2788 YINNKELNSDLTIYDDKSTLTINSFNPDPVHVGEIICKAENDAGEVSTANMITYS 2847
 QY 782 ----- 801
 Db 2848 DMFSESEAAQAEFFVGDDLDESLREEMHRTPTVM--APKFTIKIDIKAKKGHSA 2904
 QY 802 ----- 809
 Db 2905 VFECVVPDTGVCKWCKMDKGEIELIARIVQRTGPEGHITQELVLDNVTBEDAGKYTC 2964
 QY 810 ----- 836
 Db 2965 IVENTAGKTCATLVIESLEKSKKAPFIVALQDKTTKSEKVLCKVIGBPKP 3024
 QY 837 ----- 887
 Db 3025 VSWLHDNVREKNPSEKTIQESITVESGVERVTISSE-----LSHQKYTC 3075
 QY 888 ----- 910
 Db 3076 IAEENTGSKTFAFLTVOGEAPVFTKELQNKELSIGELVLSCSVKGSQPHVDVFSFE 3135
 QY 911 ----- 932
 Db 3136 TTKVETKITSRIALEDHQTWHRMVISQITKEDIVSYKAIAATNSIGTATSTSKITK 3195
 QY 933 ----- 949
 Db 3196 VEAPVPEGLKKTYSKEKEEIKMEVKVGGSPDVEWFKDDKPVSEGNHMKKNPBTGVF 3255
 QY 950 ----- 975
 Db 3256 TLVWQAAATTDAGKYTAKASNPAAGTAEAAEVTSLEKPTFVRELVTTEVKINET-- 3312
 QY 976 ----- 995
 Db 3313 --ATLSVTGVDPDSVWELKQGPVQTDSSHVIAKVEGSGSYSITIKDARLEDSGKYAC 3370
 QY 996 ----- 1027
 Db 3371 RATNP-----AGEAKTEANFAVKNLVPPEFVLEKSLPFEVKEKESITLSVKVVGTPPEPSVE 3426
 QY 1028 ----- 1041
 Db 3427 WFKDDTPISDNVHVIOQTAVGSPSLTINDARQDVGIYSCRARNEAGEALTANFGII 3486
 QY 1042 ----- 1059
 Db 3487 RDSIPPETQKLRPLREVEQETLDLKVTVIGTPVNPVWVFKDD-----KPINIDNSHIP 3540
 QY 1060 ----- 1081
 Db 1060 -----GLTTLRNGTLVAFRGH-----YFWMLS 1081

Db 3541 AKDEGSHHTL--TTQARGEDGVVYTCATNEAGEAKTTANMAVOEEIEAPLFVQGLK 3597
 QY 1082 PF--SPSPARRITEVWGISPI----- 1102
 Db 3598 PYVEQCKPAELVVRVEGPEPEVKFKGVPIAIDNQHVIEKKGENGSHTLVTKOTNNA 3657
 QY 1103 ----- 1109
 Db 3658 DFGKYTCAATNKAGKDETVCGLKIPKYSFEKQTAEEVKPLFIPLKETAPEVGDTVLEEC 3717
 QY 1110 NCEGKTF--FEKDSQYWRP----- 1126
 Db 3718 KVNKESHPOIKFPKNDQPVIEIGHQMLEVLEDGNIKLTIQNAKKEDVGAYCEAVNVAGK 3777
 QY 1127 --TN-DIK-----DAGYKPKPIF----- 1140
 Db 3778 ANINADLKIQPAKAVEHVTDESGQLEBIGQFETVGDGTASSKTDTGRGAPEFVELLSCT 3837
 QY 1141 -----KGFGGLT----- 1147
 Db 3838 VTEKQOAILCKVKGEPRPKIKWTKEGKEVMSARVRAEHKDDGTLTLTFDNVTQADAGE 3897
 QY 1148 -----CQIIVAALSTA-----KYKNWP 1163
 Db 3898 YRCEAENYGSATWTEGPIIVTLEGAPKIDGEAPDFLOPVKPAVTVGTAVLEGKISGKP 3957
 QY 1164 E-SVPFKEGGSIQ-----QY----- 1178
 Db 3958 KFSVKWYKNGEELKPSDRVKIENLDDGTORLVTNNAKLDMDEYRCEASNEFGDVWSDVT 4017
 QY 1179 IVKQEPVQKCPGRPALNYPVYGMTQVRRRPERA----- 1214
 Db 4018 LTVKEPAQVAPG-----FFKELSAIQVKETETAKFECKVSGTKPDVKWPKDGTPLKE 4069
 QY 1215 -----IGPSQTH---TIRIQYSPARLAYQKGVLUHNEVKVSI----- 1248
 Db 4070 DKRVHFESTDGTORLIVBEDSKTDQGNRYIEVS-----NDAGVANSKVPLTVVPSETL 4123
 QY 1249 -LMWGLPNV-----VTSA----- 1260
 Db 4124 KIKKGLTDVNVVTQGTKILLSVEVEGPKTVKWKYKTEITVSSQTKIVQVTESEYKLEIE 4183
 QY 1261 -----ISUPNIRK----- 1268
 Db 4184 SAEMSDTGAYRVVLSTDSFVSESSATVTVTKAAEKISLSPFKKGLADQSVKGTPLVLEV 4243
 QY 1269 -----PD----- 1270
 Db 4244 EIEGPKQVKVYKNGDEIKDGKVEDLNGKVRILTIPDFQEKDVGYSVTAANEAGEIESK 4303
 QY 1271 -----GYDYVAFSK-----DQY 1282
 Db 4304 AKVNVSAKPEIVSGLVPTTVKGETATFNVKVGPKVGVKWKYKNGKEIPDAKTONGDGS 4363
 QY 1283 YNIDVP-----SRTARAI-----TTRS 1299
 Db 4364 YSLEIPNAQVEDAADYKVVWSNDAGDADSAALTVKLADDDGKDKVKPEIVSGLIPTTVKQ 4423
 QY 1300 GQTLG---KV-----WY 1308
 Db 4424 GETATFNVKVGPKVQVKWY 4443

RESULT 11

S49915

extensin-like protein - maize

C:Species: Zea mays (maize)

C:Date: 05-Mar-1995

C:Accession: S49915

R:Rubinstein, A.L.: Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A:Description: Pex genes: pollen-specific genes with extensin-like domains.

Db 578 GYEGDCGVRRCPRDCNQGVQCDGVCTCMEGFAGEDCGLRVCPSCNCHRRGRC---ENG 633
Qy 82 R-ECD---CDAOCKYDKC---CPD---99
Db 634 RVCDSGYTGPSNATRCPADCRGRGRCVQGVCHVGYSGEDCQBEPASACPGGCGP 693
Qy 100---YESF---CAE---106
Db 694 RELCSAGQCVVEGFRGPDCAIOTCPGDCRGRGREGSCVQDGYAGEDCGEVPPIAG 753
Qy 107---VKDNKKNRKTKKTP---KPPVVDEAGS---GLDNG-136
Db 754 MRHLLUEETVTRWTRAPGNVDAYEIQFIPTTEGASPPPTARVPSSASAYDQRLAPGQ 813
Qy 137 DFXYVT---PDTSTTOHKNVSTPKIT---AKPINP-167
Db 814 EYQVTVRALRGTNWGPASKT---ITMIDGPDQLRVAVTPTTLELNMLRPQ 863
Qy 168---RPSLPNSDTSKETSILTVNKEITVETKETTINKQT---203
Db 864 AEVDRFVSVYSGNQVRLEVPSEADGTLTGLMPCVEYV---TVTAERGRAVSYP 919
Qy 204---STD---GKEKT---211
Db 920 STRANTGSSLSGLGATDEPPSPSTTQQAQAPVLQORPOELAEALRVLGKDKTGLRVA 979
Qy 212-TSAKET---QSIKTSKDLAPT---231
Db 980 WTAQPTTFHFQRLRVPRGPGNAHLLPGDVRQALVPSPPPSPYELSLRGIPPSGERS 1039
Qy 232---SKVLAKP---TPKAETTT---246
Db 1040 APLIYQIMDKDGEKGLAPLRLKLTVDVTSLSLLHWTVPGEFDSFVIQYKORD 1099
Qy 247---KGPALTTPK---EFTPTTPKE 264
Db 1100 RQVVPVEGQPSALISNDVGRKYFVLYGLVKKRHHGLVAEAKILSOTDPSVTPPR 1159
Qy 265 PASTTPKEPTTIKSAPTTPK---BPAPTTKSAPTTPKE--303
Db 1160 LGNLWVDTFPLSLHLSWTVPGQDFDSFMVQVRDRAGRQVQVVEGPDRSVIIPLDDPH 1219
Qy 304---APTTPKEPA-PTTPKEP---319
Db 1220 KYRFTLFGIANKNRHGLTADGTAPKEKEEPRHPEPPEPRLGELTVAGATADSLSLWS 1279
Qy 320---319
Db 1280 TVAQSFSFVVQYKDAQRPQAPVPTGDENEVAIPSLPDRKYMKNLYGLHGRQRVGPV 1339
Qy 320---APTTPKEAPTTPK---SAPTTPKEAPTTPK---350
Db 1340 SVWATTAQVLDTEPSTATEMEETSPTEPSTKAPESPEKPLLGELMTGSSPDSLWS 1399
Qy 351---350
Db 1400 TIPOGHFDSFTVQVRDGDQPRVMRVPQDEGVTISGLEPDHKKYKMNLYGFHGRQRVGM 1459
Qy 351---APTTPKE--PAPTTPKE-PTPT---TPKEPAPTTPKEP---383
Db 1460 SVIGVTTAEETPSPTEMEETSPTEVEETSPMEPSTEAPAPAEPLLGELTVTGSSPD 1519
Qy 384---TTPK---EP---389
Db 1520 SLSLSWTVPOGHFDSFTVQYKRGDPQVVRVRGTEETVITIGGLEFGRKYMKNLYGLHGGQ 1579
Qy 390---APTAPK---KPAPTTPKEP---405
Db 1580 RSGPVSTGVTPAQAEETPATEPPEKPRGLGELTVDTVPNSVGLWTVSGQDFDSFMVQ 1639
Qy 406---APTTPKEPAP---415

Db 1640 YKDRDGSHVVPVNAOQREATVSGLEPERKYRMNVYGLHGGQRVGLSWALTAPVPDP 1699
Qy 416 TTTKEPSPTTPKEP---APT 433
Db 1700 VT--EP---PVERLGLGELTVDTVPDSVGLSWTVAEGEFDSFLVQYKDRDQGVVPEVA 1753
Qy 434 TKSAPTTPK--EPA---445
Db 1754 TDQREVITPGLFSPSKYKFLFGIQDGRKRSTVSVEAKTVGRGDSPGAPRGLGELWTD 1813
Qy 446---445
Db 1814 PTPDSLSLSWTVPEGHFDSFVQFKORDGRPVVSVGEHRSVTISPLDSGRKYRFLVYGL 1873
Qy 446---PTTKSAPTTPKEPSPTTPKEPAP---TTPK---473
Db 1874 LGKRHRGLTTEGTETTRRAVDGAKRPSKPLGBELQVTVTSNSVGLSWTVPEGHFD 1933
Qy 474---EP 475
Db 1934 SFVIQYRDRDQGVVVPVEGSRREVSGLDPARRYKLLYGLSRDKRVGPIASIAVTEP 1993
Qy 476 APTTPKPPAPTTPKEPA---PTTPKEPAP---TTPK---506
Db 1994 APREEIKAEPAITPPASBPPLGEVTLLEAAHSLRLSWTATEGEFDSFEVQYTDENGQL 2053
Qy 507---PAPTAPK---EPAPT--518
Db 2054 QEVNVDGDQHDITISDLESDHRVLSYGFHDGQRVGPAPHIEAMTAPREDDPSPSSLST 2113
Qy 519---TPKETAP---TTPKCLTP--TTPK---537
Db 2114 TQTPSTAVPEPHIKPRLGELAVTDTTPDLSLSWTVPEGEQDFHFLYQKNGDQPKVVRV 2173
Qy 538---KLAP---TTPKEPAPTTPKEPAPTTPKEP---562
Db 2174 PGDEBVTISGLEPDHKKYKMNLYGFHNRQRMGPFVSVIGVTTABEETP---SPTMEEE 2227
Qy 563-PTTPTEP- PAPTTPKAAAPTTPKEP---587
Db 2228 TSPTEMEETSPTEBTEVPPEPEPELELTVTGSSPDSLWSLWTVPGHFDFTVQY 2287
Qy 588---GTAPTTL---624
Db 2348 PTEVEETSPTEPSTEAPPEPEPVLGELMTGSSPDSLWSLWTVPGHFDFTVQYKGR 2407
Qy 625---KEPAPTTPK 633
Db 2408 DGPQVVRVGGETEVTVBGLPEGHKKYKMNLYGLHGRVRGPVSVTAMTAREBEPASPL 2467
Qy 634 KPAPKELAPT---TTPK---647
Db 2468 KPQLGELTVDTATPDSLWSLWTVPEGEQDFHFLVQYKNGDQGVKAVRVPGEDEGVTSIGLE 2527
Qy 648---PTS---TTSK---PAPTTPKGTAPTTPKEP---672
Db 2528 PDHKKYKMNLYGFHGRQVRGPVSTVGLTVSEKQDQMTAPTDLPTAAPEPPIKPLRGLVW 2587
Qy 673---APTTPKEPAPTTPKGTAP---690
Db 2588 TDAFDSLWSLWTVPEGEQDFHFLYQKNGDQGVKAVRVPGEDEGVTSIGLEPDHKKYKMN 2647
Qy 691---TTLKE--PAPTTPK--PAPKELAPTTTKGPTSTTSKPAPTTP 730
Db 2648 YGFHGRQRMGPFVSVIGVTTABEETPGTMEQTPSPTEVEET--PGTE-VEETPSPTEP 2704
Qy 731 KETAPTTPKE---740
Db 2705 STEAPPEPEPLLGELTVTGSSPDSLWSLWTVPGHFDFTVQYKGRDQGVVVRVGGTE 2764

```
QY 741 -----PAPTPPKPAPTTPT 756
Db 2765 EVTIGGLEPGRKYMNLGLHSGQRPVSVGVTDPOEVVEETPSTPEAPEPPE 2824
QY 757 P----- 757
Db 2825 PLLGLDVTGSSPDSLSSWTVPOGHFDSFTIQYKGRDRPQVVRVGGETEVTIGLEBPR 2884
QY 758 -----PPTTSEVSTP-----TTTKEPTTIHKSPPDESTPELSAEP----- 791
Db 2895 HXYKMNLYCLHGRVRGVPVSVGVTAPOYDAMTITQTPST-----SVPEPTPKRLGVEV 2937
QY 792 -----TPKAL-----ENSPK-----EP----- 803
Db 2938 TVTDATPDSLSSWTVPOGHFDFHFLVQYKNGDQPKAVRVPDGDGVITISGLEPHKYM 2997
QY 804 -----GVPTT-----KTPAATKPEMTTAKDKTKTERDLRTTPTTTAAPKM 844
Db 2998 NLYGFHDHORVGPVSVIGVTTAAEETPSTKME-----ETPSPT 3037
QY 845 TKETATTTTETTES-----KITAT----- 863
Db 3038 MEETPSTPSTPEAPEPPEELLGELTVTGSSPDSLSSWTVPOGHFDSFTVQYKGRDGP 3097
QY 864 -----TTQVT----- 868
Db 3098 QVVRVGGETEVTVVGLEPGRKYMHLGLHGRRVGPASTVGVTSASLTTERPLAPRLGE 3157
QY 869 -----STTTQDT-----PF-----KITLTKTTLAP----- 890
Db 3158 LAVAVTSDTARLSWTVGQGFDSFLVQYKDVQGPQVAVAGDLREVTVSSLAPGRKYK 3217
QY 891 -----KVTTTKTITTTTEIMNKPEETAKP----- 914
Db 3218 FLFLGLRDEKHGFPVSADAKTLPTD-----KPAPRLGELTVDTPGSGLSWT 3267
QY 915 -----KDR-----A 918
Db 3268 PEGFDSFMVQYKDRDQRPVFAADQREVTVGLEPNRKYKFLYGLVGRKRLGPISA 3327
QY 919 TNSKATTPKQKP----- 931
Db 3328 EGSTAPLEKERQPPRLGELTVDTEPNSRLSWTVAGRFDSFVQYRGTDQGRMPV 3387
QY 932 -----TKAPKXP 938
Db 3388 AADQREFTVGLRPGKRYKFLYGLLGGQRLGPASVLGMTAPEDTTPAPWHAATEAPKPP 3447
QY 939 ----- 938
Db 3448 EGRPLGLAVRDSFDSLRLSSWSVVGPFDSFVVQYQDTGQPOALLVGDNQKVLVSL 3507
QY 939 ----- 938
Db 3508 EPSTSYEFFLYGLHEGRKLGFPVSAEGTGPVAGQTPGEGPRLSHLSLTDVTTSSLRLN 3567
QY 939 -----TSTKPKMPRVKPKTTPTPPKMTSTWPELNP----- 971
Db 3568 WEAPPEAFDSFLLRFGVPSPSTLPQLRPLQLRELTPVGTTR-SAVLRDLHPGLTYLTL 3626
QY 972 -----TSRAEAML-----QITTRNQTPNSKLVE---VNPKE 1002
Db 3627 YLGRGPHKADSIQGTARTLSPVLSRDLQSEIRETSARVSWTPPTSRVDGFKVSVQLA 3686
QY 1003 DAG-----GAGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIIN 1043
Db 3687 DGGPQSVQVQDRTQKLEGLLPGAQYEVTVVSVRG---FESEPLTGLTITVPD-----G 3738
QY 1044 PM-----LSDETINCNK-----PVDGLTILR 1065
Db 3739 PTHLRALNLTDESALLHWKPPQTPVDYDVKVTAFGAPFSLOASAPGSADVDPLOGLVHT 3798
```

```
QY 1066 N--GTLVAFRGHYF-----WMLSPSPSP----- 1086
Db 3799 NYTATLRLGRGNFTSPASITFTTGLEAPQDLEAKVETPTALTWTAPESPTGYLLSF 3858
QY 1087 -SPARRITEV-----WG-----IPSP 1101
Db 3859 NTPGGQTQEIILLPGGVTSQHLRGLPFPSTYTWLRAMMGDSFTPPVSVSTFTTGGURIFPP 3918
QY 1102 ID-----TVFTRCNCE-----GKTFPFKQSQ 1122
Db 3919 RDCGEMONGVSTSTFTTIFLNGRERPLNVFCMDTGGGWLVFQRRMDGKTDFWRD-- 3976
QY 1123 YWRFTNDIKDAGYKPIFKGFGGLTGQIVAAALSTAKYKNW--PESVYFKRGSGSQIYIY 1180
Db 3977 -W-----EDYAH-----GFGNISGEF-----WLGNEALHSLTKAGD----- 4006
QY 1181 KOEPVQKCPRRPALNYPVYGEMTQVRRRFRPERRAIGPSQTHIRIQYSPARLAYQDKGVL 1240
Db 4007 -----YSLRVD-----LRAGDEAVF 4021
QY 1241 HNEKVSVILWRLGNVWVTSALSLNIRKPDGYD--YVAFSKDQYNNIDVPSRTARA---IT 1296
Db 4022 AQ-----YDSFQVDSADVEYRLHLEGYHGTAGDSMS 4052
QY 1297 TRSGQTLS-----KVWY-NC 1310
Db 4053 YHSGSVFSARDRPNLLISCAVSYRGAWWYRNC 4086

RESULT 13
A41819
proline-rich peptides 637K precursor, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 09-Jul-2004
C:Accession: A42663; A41819; A31966; B20593; A20593
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the
le exon.
A:Reference number: A42663; MUID:92250652; PMID:1577819
A:Accession: A42663
A:Molecule type: DNA; mRNA
A:Residues: 1-3706, 'I', 3708-4077, 'F', 4079-4155, 'S', 4157-5761 <DE2>
A:Cross-references: UNIPROT:Q63455; GB:M86514
A:Experimental source: ventral prostate
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBI:100347, NCBIP:100348)
R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
submitted to the Protein Sequence Database, April 1992
A:Reference number: A41819
A:Accession: A41819
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-5762 <DE1>
R:Hemschoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Win
J. Biol. Chem. 263, 19159-19165, 1988
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich p
A:Reference number: A31966; MUID:89066721; PMID:3198617
A:Accession: A31966
A:Molecule type: mRNA
A:Residues: 3372-3540 <HEM>
A:Cross-references: GB:M20721; GB:J04188; NID:G206397; PIDN:AAA41950.1; PID:G554494
R:Peeters, B.; Heyns, W.; Bossyrs, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A:Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary
A:Reference number: A94675; MUID:84061859; PMID:6685733
A:Accession: B20593
A:Molecule type: protein
A:Residues: 2020-2057 <PRE>
A>Note: this peptide, designated proline-rich polypeptide V, can be found at several lo
A:Accession: A20593
A:Molecule type: protein
A:Residues: 2822-2859 <PE2>
```

A>Note: this peptide, designated proline-rich polypeptide IV, can be found at several loci
 C;Genetics:
 A;Introns: #status absent
 A;Note: single copy gene with no introns
 C;Superfamily: rat prostatic proline-rich peptides 637K precursor
 C;Keywords: prostate; tandem repeat
 F;1-26/Domain: signal sequence #status predicted <SIG>
 F;27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <MA>

Query Match 12.5%; Score 885.5; DB 2; Length 5762;
 Best Local Similarity 10.6%; Pred. No. 6.9e-10;
 Matches 533; Conservative 172; Mismatches 447; Indels 3865; Gaps 151;

| | | | | |
|----|------|--|-------------------------------------|-----|
| QY | 4 | KTLPIYLLLLSVFVIQQVSSQDLLSCAGRCGEGYRDATCNCYDNCQ | -----OPTQNEYIGQLPNVTVRPVDVA | 486 |
| Db | 456 | ETLPIHEI | ----- | 51 |
| QY | 52 | -----HVMCCPDPRVCTAELSCKRCFESFERGRE | ----- | 83 |
| Db | 487 | LTVTSEPVKETESFLAPQEPFVALEYSD | -----VEPFVNEEPPVQAP | 531 |
| QY | 84 | -----CCDCAOCKYDKCCPDYESFCAEVKDNKNRTKKP | ----- | 119 |
| Db | 532 | ETPGESOFESQLEVPAGATEYD | -----EBF-----KTSATEQEQLAQFPENDEVTVL | 578 |
| QY | 120 | -----TPKP-----PVDKAGSGLDNGDF | ----- | 138 |
| Db | 579 | PSNHQAQHSILSNVTQDPLDLITITEKPMEMGTSPVYDAAAAPVEVFLSDQQGVLS | 638 | |
| QY | 139 | -----KVTPDNTSTQH | ----- | 150 |
| Db | 639 | QSLPILYDLSQOQYTTGISQSEGEPPTQETPEHSGVMHTERVAQPPGHVETGP | 698 | |
| QY | 151 | -----NKVSTSPKITTAKP | -----INRPSLPNSD | 176 |
| Db | 699 | PLGHGQVHPALQNIITQYSTPPEKQHSVGLGVFCHLEEFVSVPSPQOENSAMHSVT | 758 | |
| QY | 177 | -----TSKETSLTVNKETVE | ----- | 192 |
| Db | 759 | DMFLSPVDLQTIFRSTQSKSYKTIKHEDLALTIITPEPSLEDGSIILFPQEDLLQPIDSTG | 818 | |
| QY | 193 | -----TK-----EFTTNTKQTS | -----TDGK | 208 |
| Db | 819 | QGEFHSIKTSLSKPPYVNTKSSAFQETMSTYNSKQVLDSSHLTKTSELPPDYTMGL | 878 | |
| QY | 209 | EKTSIAKETQS | -----LEKTSAKDLAPTSKVLAKPT | 239 |
| Db | 879 | EPSLYQOITQSSPKSLHENPKSPFVLKPSAAQALKPRKEKLSPTNNMIPHHSKPLKNM | 938 | |
| QY | 240 | ----- | ----- | 239 |
| Db | 939 | VTHIPAHKMTVPRQIQEQGEYTISSNGSQPLDLEVLITSLGIIPEVKHILPKRTVNPQT | 998 | |
| QY | 240 | -----PRAETTKGP-----ALTTPE | ----- | 256 |
| Db | 999 | YSQVKISHSQHVETQHPNSSETTVQPLDLFAINLQTPKENFAQTQFDQTTQTMIGPPKE | 1058 | |
| QY | 257 | ----- | -----PTPTTPKEPAST | 268 |
| Db | 1059 | VIAQAPHEHGTIPIQDQAEYSTLPTVSVFQPLDQELTITSEATREPPHPTVPOQTIIIV | 1118 | |
| QY | 269 | TPKE | -----PTPT-----TIKSAPTPKE | 287 |
| Db | 1119 | HPPEHPLVHSBQVYVQHNPTEAIQPLDLELTITPQTAEGELPQTLQDSTQIIEPP | 1178 | |
| QY | 288 | ----- | -----PAPT----- | 291 |
| Db | 1179 | TVVGVPIYIEVTVQTTSDQAEYPPSPVTSFQSLDLELTITPQTAEGELPQTLQDSTQIIEPP | 1238 | |
| QY | 292 | ----- | ----- | 291 |
| Db | 1239 | IVNPPHPLVHSEQVHTQHNLTEATVQPLDPLDITITPQPTTEGELPQTLQDSTQIIII | 1298 | |

| | | | | |
|----|------|--|--------------------|-----|
| QY | 292 | -----TTKSA----- | -----PTTPKE----- | 302 |
| Db | 1299 | EPPIKWALVYQEVSDQAEYTTSTSVSFQPLDQELTITSEAIRPHHPTVPOQTIV | 1358 | |
| QY | 303 | -----PAPT-----TTKEPAPTTPKE | ----- | 318 |
| Db | 1359 | HPTKHLPLVHSEQTHQHNPTVTVQPLDLELTITMTPOPTAEGELPQTLQDSTQIIEPPTV | 1418 | |
| QY | 319 | -----PAPT----- | ----- | 323 |
| Db | 1419 | VVGVPYIEVTVQTTSDQAEYPPSPVTSFQSLDQELTITPTEPTEHPTQKTTVFPDPMY | 1478 | |
| QY | 324 | -----TKEPAP-----TTKSAPTTTPKEP | -----A | 343 |
| Db | 1479 | TDVTLPOQVSVQHLKPTGIVQPLDLELTITPQPTPEGELSQTVQEBSTTKQKPHKEVVA | 1538 | |
| QY | 344 | PT-----TPKK----- | ----- | 349 |
| Db | 1539 | PVPYQAVTVPTPSQQAQYQKSLQPLDLELTITVTSEPTKEAYHSTISKNLSAINPQYVHI | 1598 | |
| QY | 350 | -----PAPT----- | -----PKE----- | 357 |
| Db | 1599 | QHNPAAEATVQPLDLELTITSSLSQPTAEGELLYSMQETVQISEPPKQVTVTPVEYQEA | 1658 | |
| QY | 358 | -PAP----- | -----TTPKEPTTPTPKE | 373 |
| Db | 1659 | VPAPVQDQAKYPLSSIVLSNLSDQELTISELLGEAHLTPDPTWVLPKDKQGIYPDH | 1718 | |
| QY | 374 | ----- | ----- | 373 |
| Db | 1719 | DHKKHLNLTETVNTQPHLEHTVQHPTIEERSQSIQKKTQITEFGKVVPLAQESEV | 1778 | |
| QY | 374 | -PAPTTKBPAPTTP | ----- | 386 |
| Db | 1779 | TIPMILKETAPTTPHSMALQSLDEKLTIHSHSPGWTQHANKESKGTHTGKILLDYAE | 1838 | |
| QY | 387 | ----- | -----KE | 388 |
| Db | 1839 | PNMEIELKHGLFLKTTTEATTESENTNQMTSKLKQVTLFTQNKKSMLPALVESQESQ | 1898 | |
| QY | 389 | PAPTAPKPK-----APTPK-----BPAPT | ----- | 408 |
| Db | 1899 | PPNNMSLQPLDQELTSSQPHGWVPHIPNTPBKIYLHYAEP-PTGPFVPEPPELFLKTTK | 1957 | |
| QY | 409 | -----TPKE-----PAP----- | -----TTTKE | 420 |
| Db | 1958 | SRPVQGTATQMAASPKEMVSRAPENKEAVLSQGEDQDESPPPNMSLOSLOQELTSSQ | 2017 | |
| QY | 421 | PSPTTPKEP-----APT-----TTKSAP | ----- | 438 |
| Db | 2018 | PHGWIPHPNTHGKIYLHYAEPPTGPFVEPPDLFFLTKTKSKPEVWTLTRTDKSRKEMVS | 2077 | |
| QY | 439 | -----TTTKEP----- | ----- | 444 |
| Db | 2078 | QSPKYEEAVLPVHGQGEESRPPNNMSLOSLEQELTSSQPHGWVPHPPNTHGKIYLHYA | 2137 | |
| QY | 445 | -APT-----TTKSAP----- | ----- | 453 |
| Db | 2138 | EPPTGPFVEPPDLFFLTKTKSKPVQGTATRWKSPSEMVSLDPENKEAVFPAQGEKGES | 2197 | |
| QY | 454 | ----- | -----TTP | 456 |
| Db | 2198 | PSSPNMSLOSLEHFMSSQPHGWIHPHPPKTPDKIYLHYAEPPTGPFVEPPDLFLRTTK | 2257 | |
| QY | 457 | KEPSTTTTKBPAPTTPKEPAPTTP-----KCAP----- | -----T | 486 |
| Db | 2258 | SKPVQCHITTEMA-KSPKEMVSTPEYKEAVLSGCGEDQDESPPNTSLKSLDQEVAMS | 2316 | |
| QY | 487 | TPKEPAPTTPK-----EPAP-----TTTK-KPAPTAP | -----KEP | 515 |
| Db | 2317 | QPHSGVHPHPPKTPGKIYLHSTPEPPGPFVKPTDILLVKTTKSKPAEWTPPRIDKLLKEM | 2376 | |
| QY | 516 | APTPK-ETA----- | -----PTPKKL | 531 |

QY 746 PKKPAPTPE--TPPPTTSE-VSTPTTTEKPTTIHKSPDBESTPSELSAETPPKALENSPK 802
 Db 802 TTTTEKTKTKTEKTTKTESATETTTSEPTST-----TEST----- 838
 QY 803 PGVPTTKTAATKPEMTTAKDKTTRDURLTTPETTTA----- 840
 Db 839 ----TVDTSATTEESSTAETTTSAE---TSETTTSAAITGESPENTALQSSSQ 891
 QY 841 -----APMTKETATTTTEKTTESKITATTTQ-VTST---TTQDTTPFK 879
 Db 892 SEENESSAEPKPGARRDFVFKKHTTVKPAETTSAGVAASITTTTPTTEKSTTLETTPIE 951
 QY 880 ITTLKTTTLAPKVT---TTKTKTTTTEIMKKPETAAPKPKDRATNSKATPKPQ----- 929
 Db 952 ATTLNEVTGFAVIGAPVDETTINTLELUSK-----INNQTQSQPKPTDISKTD 1001
 QY 930 -----KPTKAPKKTSTKKPTMPRVKPKTPTTPKMTSTMPELN----- 970
 Db 1002 LSSLISGLIGSFTKAPWAPI-----HITDAAAFVTATEASLNDGSKKIID 1048
 QY 971 ---PSTRIAEAMLOTTTRENQTSKLVEVNPKSESDAGAGETPHMLLRPHVFMPEVTP 1027
 Db 1049 EAQPTDEIRRA-----OPTN-----EMDK 1067
 QY 1028 DMDDLPRVNOGIII-----NPMI----- 1046
 Db 1068 EMEPEKRIEORIQMQAKRLREBELLEKQLOQEIEKARNEMIERKQMKLOOLBELKE 1127
 QY 1047 -----SDETNINCNGKPDVGLTTLRNGTLVAFRGHYFWMLSP 1082
 Db 1128 ABERQORVLEQERLOEQERQLAEKEAELAFG-----SISTTTEASKFKYR----- 1174
 QY 1083 FSPSPAR-----RITFVWGPISPIDTVFTRCNCE-GKTFPFKDSQWFRFTNDIKAGYP 1136
 Db 1175 ---LRPAQCAAINKFRVNTIDPSEWI---QKNCEFAKRYF-----PEASCP 1216
 QY 1137 KPIFKFGGLTGQIVAAALSTAKYKNWPESVYFF 1169
 Db 1217 -----QIQALI-----ESCFAP 1228

RESULT 15
 A35175
 N:Title: mucin 1 precursor, repetitive splice form A [validated] - human
 N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episialin
 N:Contains: mucin; polymorphic epithelial mucin (PEM)
 C:Species: Homo sapiens (man)
 C>Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text change 02-Jun-2000
 C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51
 R:Lightenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilken, J.
 J. Biol. Chem. 265, 5573-5578, 1990
 A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene and
 A:Reference number: A35175; MUID:9020794; PMID:2318825
 A:Accession: A35175
 A:Molecule type: mRNA
 A:Residues: 1-952,1033-1344 <LIG1>
 A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124; GB:
 A:Experimental source: Splice form A
 A:Note: GenBank entries H0MEPISIA1 and H0MEPISIA2 present only the amino-and carboxyl-en
 A:Accession: B35175
 A:Molecule type: mRNA
 A:Residues: 1-19,29-952,1033-1344 <LIG2>
 A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129; GB:
 A:Experimental source: splice form B
 A:Note: GenBank entries H0MEPISIB1 and H0MEPISIB2 present only the amino-and carboxyl-en
 R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel
 J. Biol. Chem. 265, 15286-15293, 1990
 A:Title: Molecular cloning and expression of human tumor-associated polymorphic epitheli
 A:Reference number: A35886; MUID:90368715; PMID:1697589
 A:Accession: A35886
 A:Status: not compared with conceptual translation

A:Molecule type: mRNA
 A:Residues: 1-19,29-992,1033-1344 <GEN>
 A:Cross-references: GB:J05581; NID:g188869; PIDN:AAAS9876.1; PID:g188870
 A:Note: GenBank entry HUMUCAB includes one copy of the tandemly repeated sequence
 J:Ian, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
 J. Biol. Chem. 265, 15294-15299, 1990
 A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
 A:Reference number: A35887; MUID:90368716; PMID:2394722
 A:Accession: A35887
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
 A:Cross-references: GB:J05581; NID:g189598; PIDN:AAA60019.1; PID:g189599
 A:Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated sequence
 R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.;
 Eur. J. Biochem. 189, 463-473, 1990
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
 A:Reference number: S10571; MUID:90276413; PMID:2351132
 A:Accession: S10572
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRB>
 A:Cross-references: EMBL:X52229; NID:g37053
 R:Wreschner, D.H.
 submitted to the EMBL Data Library, March 1990
 A:Reference number: S40293
 A:Accession: S40293
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <WR2>
 A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
 R:Abe, M.; Siddiqui, J.; Kufe, D.
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
 A:Reference number: A36735; MUID:90088473; PMID:2597151
 A:Accession: A36735
 A:Molecule type: mRNA
 A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
 A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
 R:Masuzawa, Y.; Miyauchi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.
 J. Biochem. 112, 609-615, 1992
 A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglut
 A:Reference number: JX0235; MUID:93123189; PMID:1478919
 A:Accession: PX0066
 A:Molecule type: mRNA
 A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
 A:Experimental source: Gastric carcinoma cell
 R:Zrian-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
 FEBS Lett. 336, 130-136, 1994
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine r
 A:Reference number: S51026; MUID:95080414; PMID:7988707
 A:Contents: annotation
 A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region an
 C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
 Partial repeats. The repeat shown is defined by SmaI nuclease sites.
 C:Comment: Serine and threonine residues in the tandem repeat domain are extensively glycy
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
 C:Genetics:
 A:Gene: GDB:MUC1; PUM
 A:Cross-references: GDB:120705; OMIM:158340
 A:Map position: 1q21-1q23
 A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
 A:Superfamily: polymorphic epithelial mucin
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
 F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
 F:1-62/Region: mucin 1 amino-terminal non-repetitive
 F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>
 F:1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
 F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
 F:1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice form
 F:138-1017/Region: 20-residue repeats (GSTAPPAGHGVTSAPDTPAP)
 F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
 F:1245-1272/Domain: transmembrane #status predicted <IRM>
 F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 12.3%; Score 867.5; DB 1; Length 1344;
 Best Local Similarity 21.8%; Pred. No. 7.1e-11;
 Matches 371; Conservative 93; Mismatches 423; Indels 815; Gaps 84;

QY 7 PIVLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCYHQHYMECCPDFREVCTA 66
 DB 8 PFLLLLLVLL----- 18

QY 67 ELSCKGRCFESPERGECDDAQCKKYDKCCPDYESFCAEVKDNKKNRTKKKPTPKPPVV 126
 DB 19 ----- 18

QY 127 DEAGSLDNGDKVITPDTSTTQHNKVTSPKLTITAKPINPRSLPPNSDTSKETSILTVN 186
 DB 19 -----TATTAP 24

QY 187 KETTVETKETTITNKOTSDGKEKTTSAKETOSIEKTSAKD-LAPTSKVLAKPTP-KAET 244
 DB 25 KQATV-----VTGSCHASSTPGGEKETSATORSVSPSTEKNAVMTSSVLSSHSPGSGS 80

QY 245 TTKGPALTTKEPTPTTPKEPA-----STTP-----KE 272
 DB 81 TTQGGQDVL-----APAT-----EPASGSAATWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 273 PTPPT-----LKSAPTTKPEPAPTT-----KSAPTTKPEPAPTTKPE-----A 312
 DB 135 PAPGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAP 190

QY 313 PTTKPEPAPTTKPEPAPTTKSAPTTKPEPAPTT-----KKPAPTTKPE----- 358
 DB 191 PGT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGV 247

QY 359 --APTTPKEPTTP-----KEPAPTTKPE-----APTTPKEPAPTTAP----- 394
 DB 248 TGAOTRPAAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGV 307

QY 395 -----KKPAPTTKPE-----APTTPKEPAPTTKPE-----SPTTKPEPAPTTK----- 435
 DB 308 TSAPOTRPAAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPA 363

QY 436 ----SAPTTTKPEPAPTT-----KSAPTTKPEPSPTTKPE-----APTTPKEPAP 477
 DB 364 AHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 419

QY 478 TTP-----KKPAP--TTP-----KEPAPTTKPE-----APTTPKKEPAP 509
 DB 420 TAPPAGHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 479

QY 510 TAPKEPAPTTKETAPTTPKLTPTTPEKLAPTTPKEPAPTTPE-----ELAPTTPPEPTP 565
 DB 480 TAPPAGHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 539

QY 566 TTP-----EEPAP--TTPKA-----AAPTTPKEPAPTT-----KEPAPTT 599
 DB 540 TAPPAGHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 599

QY 600 PKEP-----APT-----PKETAP-----TTPKG-----TAPTTLKEPAP 629
 DB 600 TAPPAGHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 659

QY 630 TTP-----KKPAPKELAPTTKPE-----TSITSDKAP--TTPKG-----TAPTTP 669
 DB 660 TAPPAGHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGS 714

QY 670 KPEPAPTT-----KEPAP--TTPKG-----TAPTTLKEPAPTT-----KK 703
 DB 715 PAPGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT 774

QY 704 PAPKELAPTTKGTSTSDKAP--TTPK-----ETAPTTPKEPAPTT-----K 747
 DB 775 PAPGSTAP--PAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT--RPAGSTAPPAHGVTSAPDT 833

QY 748 KPAPTTPTPTTSEVSTPTTTTKEPTT-----IHKSPDESTPLSAPETPKALENSP 800
 DB 834 REAPGS--TAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP 891

QY 801 KEPGVPTTKTPAA-----TKPEMTTAKOK--TTERDLRTTPTT-----TAAP 842
 DB 892 DTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTSAP 951

QY 843 KMTKETATTTETTESKITATTQTQTTTQTTQTTTQTTTQTTTQTTTQTTTQTTTQTTT 902
 DB 952 -----DTRP----- 955

QY 903 EIMNKPEETAQKDRATNSKATTPKEQKTKAPKPTSTKKKTMVRVKPKTTPTPRKM 962
 DB 956 ----APGSTAPPAHGVTSAPDTRPAP--GSTAPPAHGVTSAPDTRP--APGSTAPPAHG 1006

QY 963 TSTMDELNP-----TSRIAEAMLQTTTRPNOTP--NSKLVEVNPKSDAGAGETPHMLL 1016
 DB 1007 VTSAPDTRPAPGSTAPPAHGV--TSAPDNRRPALGSTAPVHNVTISASGSAG----- 1056

QY 1017 RPHVFMPEVTPMDXLPVFNQGIINPMLSDETNICNGKPVGDLTTLRNGTLVAFRGHY 1076
 DB 1057 -----SASTLVHNG-----TSARATTTPASKS-- 1078

QY 1077 FWMLSPEPSPSPARRITEVWGIPSPIDTVTRCCEGKT----- 1115
 DB 1079 ----TPFSIPS-----HHSDDPTTLASHSTKTKDASSTHHSTVPEPLTSSNHSTS 1122

QY 1116 ----PFF-----KDSQVWRETDNDKAGYKPIFK--GFGGLT- 1147
 DB 1123 POLSTGVSTFFFLSHISNLQFNSSLEDPSTDIYQELQDISFWL--QIYKQGGFLGLSN 1180

QY 1148 ----GQIVVAULSTAKYKNWPESVFFKRG--STQOYIYKQEPVQK----- 1187
 DB 1181 IKFRPGSVVVQULTA-----FREGTINVHDVETQFNQYKTEAASRYNLTISDVS 1229

QY 1188 -----CPGRR----- 1192
 DB 1230 VSDVFFPFSAGSAGVPGWGIALLVLCVLVALIIVYLIALAVCCRRKNYGOLDIFPAR 1289

QY 1193 ----PALNYPVYGEVTQVRRRRFRFAIGESQTHIRIQYSP-----ARLAYQDKG 1238
 DB 1290 DTYHPMSEYPTY-----HTHGRVVPSSST-----DRSPYEKVSAGNGSSLSYTNPA 1336

QY 1239 VLNHEVVKVGSILWRGLPNVVTSA 1260
 DB 1337 V-----AATSA 1342

Search completed: October 13, 2004, 12:03:35
 Job time : 106.704 secs

This Page Blank (uspto)